OXFORD

modern optics

SECOND EDITION

B. D. GUENTHER

MODERN OPTICS

Modern Optics

Second Edition

B. D. Guenther

Duke University, USA

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published by Wiley in 1990
Second Edition published in 2015

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2014957917

ISBN 978–0–19–873877–0

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

The second edition of this textbook is designed for use in a standard physics course on optics. The book is the result of a one-semester elective course that has been taught to juniors, seniors, and first-year graduate students in physics and engineering at Duke University for 13 years and has been used as a graduate text in a number of universities. Students who take this course should have completed an introductory physics course and mathematics courses through differential equations. Electricity and magnetism can be taken concurrently.

Modern Optics differs from the classical approach of most textbooks on this subject in that its treatment of optics includes some material that is not found in more conventional textbooks. These topics include nonlinear optics, guided waves, Gaussian and Bessel beams, photonic structures, surface plasmons, imaging and computational imaging, and light modulation. Moreover, a selection of optional material is provided for the instructor so that the course content can reflect the interest of the instructor and the students. Basic derivations are included to make the book appealing to physics departments, and design concepts are included to make the book appealing to engineering departments. Because of the material covered here, the electrical engineering and biomedical engineering departments at Duke have made the corresponding optics course a prerequisite for some of their advanced courses in optical communications and medical imaging.

Before the 1960s, the only contact that the average person had with optics was a camera lens or eyeglasses. Geometrical optics was quite adequate for the design of these systems, and it was natural to emphasize this aspect of optics in a curriculum. The approach used introduced the students to the theory and to examples of the application of the theory, accomplished by a description of a large variety of optical instruments. The reason for this approach was that lens design is found to be quite tedious, and the optimization of a lens design is more easily described than accomplished.

Today, the student is exposed to many more optical systems. Everyone encounters supermarket scanners, copying machines, compact disk players, holograms, and discussions of fiber-optic communications. In the research environment, lasers, optical modulators, fiber-optic interconnects, and nonlinear optics have become important. Upon graduation, many students will be called on to participate in the use or development of these modern optical systems. An elementary discussion of geometrical optics and a review of classical optical instruments will not adequately prepare the student for these demands.

The second edition retains the emphasis on both fundamental principles of optics and an exposure to actual optics engineering problems and solutions. A large portion of the conventional treatment of geometrical optics and a discussion of classical optical systems is not included. In their place, a geometrical optics discussion of fiber optics has been included and, in this second edition, a discussion of photonic structures with holography being considered as a special case of photonic structures. Rather than describe a host of optical systems, a few systems, such as the Fabry–Perot interferometer, are examined using a variety of theories. The second edition retains its emphasis on diffraction and the use of Fourier theory to describe the operation of an optical system.

To allow the development of a one-year course in modern optics, a number of topics have been added or expanded. The classical theory of the interaction of light and materials has been expanded to include surface plasmons. A discussion of electro-optic and magneto-optic effects is used to introduce optical modulators, and a discussion of nonlinear optics is constructed around second-harmonic generation. Because of the importance of birefringence in optical modulators and nonlinear optics, an expanded discussion of optical anisotropy has been included. This is a departure from most texts, which ignore anisotropy because of the need to use tensors. In modern optics, anisotropy is an important design tool, and its treatment allows a discussion of the design of optical modulators and phase matching in nonlinear materials.

The first two chapters review wave theory and electromagnetic theory. Except for the section on polarization in Chapter 2, these chapters could be used as reading assignments for well-prepared students. Chapter 3 discusses reflection and refraction and utilizes the boundary conditions of Maxwell's equations to obtain the fraction of light reflected and refracted at a surface.

Chapter 4 discusses interference of waves and describes several instruments that are used to measure interference. Two of the interferometers—Young's two-slit experiment and the Fabry–Perot interferometer—should receive emphasis in discussions of this chapter because of the role they play in later discussions. An appendix to this chapter provides a brief introduction to some of the design techniques that are used to produce multilayer interference filters. All of the appendices in the book are included to fill in gaps in students' knowledge and to provide some flexibility for the instructor. The appendices may therefore be ignored or used as the subject matter for special assignments.

The treatment of geometrical optics, presented in Chapter 5, is not traditional. It was through the reduction of traditional subject matter that space has been freed to allow the introduction of more modern topics. A brief introduction to the matrix formalism used in lens design is presented, and its use is demonstrated by analyzing a confocal Fabry–Perot resonator. Geometrical optics and the concept of interference are used to analyze the propagation of light in a fiber. This introduction to fiber optics is then extended through the use of the Lagrangian formulation to propagation in a graded-index optical fiber. The connection between the matrix equations and the more familiar lens equations is established in Appendix 5A. Because of their importance in the Graduate Records Exam, aberrations are treated in Appendix 5B.

A formal connection between geometrical and wave optics was presented in Chapter 5 in the first edition. It has been removed from this edition because it was not found useful in the classroom.

The Fourier theory in Chapter 6 is presented as a review of and refresher on the subject. It is an important element in the discussions of the concept of coherence in Chapter 8 and of Fraunhofer diffraction in Chapter 10. The discussions of optical signal processing in Appendix 10A and of imaging in Appendix 10B draw heavily on Fourier theory. The treatment of imaging has been expanded to include computational imaging and coded apertures.

The discussion of dispersion given in Chapter 7 could be delayed and combined with the other chapters on material interactions (Chapters 13 and 15). It is included here to justify the discussion of coherence in Chapter 8. The discussion of dispersion in materials had as its objective the development by the student of a unifying view of the interaction of light and matter. It has been expanded to include surface plasmons.

The development of coherence theory in Chapter 8 is built around applications of the theory to spectroscopy and astronomy. It is a very difficult subject, but building the theory around the methods used to measure coherence should make the subject more intelligible.

Both the Fresnel and Gaussian wave formalisms of diffraction are introduced in Chapter 9. The Gaussian wave formalism is used to analyze a Fabry–Perot cavity and thin lens. This chapter can be skipped, and the material introducing the Fresnel–Huygens integral can be covered in a single lecture.

The Fresnel formalism is expanded and discussed in Chapters 10 and 11. Fraunhofer diffraction is treated from a linear system viewpoint in Chapter 10, and applications of the theory to signal processing and imaging are presented in Appendices 10A and 10B. These two appendices are the most important in the book. Fresnel diffraction is introduced in Chapter 11, where it is used to interpret Fermat's principle and analyze zone plates and pinhole cameras.

In this second edition, Chapter 12 is constructed around a presentation of photonic structures. Guided by the presentation found in the book *Photonic Crystals: Molding the Flow of Light* by Joannopoulos, Johnson, Winn, and Meade (Princeton University Press, Second Edition, 2008), an overview of diffractive structures and their applications is given. Multidimensional periodic modulation of the index of refraction is shown to produce light propagation properties similar to the behavior of electrons in crystals. The similarity to holograms is discussed. A simple theory of holography is presented, and Fresnel theory is used to discuss the operation of a hologram. To generate the space needed to discuss photonic structures, much of the material on holography found in the first edition has been removed.

Chapter 13 uses the introduction of polarizers and retarders as a basis for the development of the theory of the propagation of light in anisotropic materials. The treatment of anisotropic materials is expanded over the conventional presentation to allow an easy transition into the discussion of light modulation in Chapter 14. The many geometrical constructions used in the discussion of anisotropy are confusing to everyone. To try to make the material understandable, only one construction is used in Chapter 13. To provide the student with reference material to aid in reading other books and papers, the other constructions are discussed in the appendices.

The discussion of modulators in Chapter 14 provides an application-based introduction to electro- and magneto-optic interactions. The design of an electro-optic modulator provides the student with an example of the use of tensors. The material interactions presented in Chapters 14 and 15 require the use of tensors, a subject normally avoided in an undergraduate curriculum. Tensor notation has been used in this book because it is key in the understanding of many optical devices. Some familiarity with tensors removes much of the "magic" associated with the design of modulators and the applications of phase matching discussed in Chapter 15.

The subject of nonlinear optics in Chapter 15 is developed by using examples based on frequency doubling. Only a few brief comments are made about third-order nonlinearities. The additional discussion of third-order processes is best presented from a quantum mechanical viewpoint. It was thought that this would be best done in a separate course. The material presented in this chapter should prepare the student to immediately undertake a course in nonlinear optics.

Enough material has been included for a one-year course in optics. Chapters 2–4, 6, 9 (excluding Gaussian waves), and 10 contain the core material and could be used in a one-quarter course. By adding Chapters 7 and 8 along with Appendices 10A and 10B, a one-semester course can be created. The instructor can alter the subjects discussed from year to year by adding topics such as Appendix 4A, the guided wave discussion of Chapter 5, or the discussion of holography in Chapter 12 in place of Appendices 10A and 10B. A less demanding one-semester course can be created by ignoring Chapter 8 and substituting Chapter 5, Appendix 4A, or possibly Chapter 12. In anticipation of developing skill and knowledge, the subject matter and problems increase in difficulty as the student moves through the book.

A number of people provided help in the preparation of this book. Those who provided photographs or drawings are identified in the figure captions. Their generosity is most appreciated. Dr. Frank DeLucia provided the initial motivation for writing the book. Many ideas and concepts are the result of breakfast discussions with A. VanderLugt. The book would never have gone past the note stage without the equation writer written by Dennis Venable. Duncan Steel reviewed some of the additions to the second edition, and I want to thank him for that effort. Thomas Stone provided ideas, photographic skills, and encouragement during the preparation of most of the photographs in this book. His enthusiasm kept me working.

Very special thanks must go to Nicholas George. He loaned me equipment and laboratory space to prepare many of the photographs. His encouragement prevented me from shelving the project, and his technical discussions provided me with an improved understanding of optics.

ROBERT D. GUENTHER

Contents

1 Wave Theory 1

1.1 Introduction 1
1.2 Traveling Waves 1
1.3 Wave Equation 4
1.4 Transmission of Energy 5
1.5 Three Dimensions 6
1.6 Attenuation of Waves 10
1.7 Summary 11
1.8 Problems 12

2 Electromagnetic Theory 15

2.1 Introduction 15
2.2 Maxwell's Equations 16
2.2.1 Gauss's Law 16
2.2.1.1 Gauss's (Coulomb's) Law for the Electric Field 16
2.2.1.2 Gauss's Law for the Magnetic Field 17
2.2.2 Faraday's Law 17
2.2.3 Ampére's Law (Law of Biot and Savart) 17
2.2.4 Constitutive Relations 18
2.3 Free Space 19
2.4 Wave Equation 19
2.5 Transverse Waves 21
2.6 Interdependence of **E** and **B** 22
2.7 Energy Density and Flow 24
2.8 Polarization 28
2.8.1 Polarization Ellipse 29
2.8.1.1 Linear Polarization 32
2.8.1.2 Circular Polarization 33
2.8.2 Stokes Parameters 36
2.8.3 Jones Vector 39
2.9 Propagation in a Conducting Medium 40
2.10 Summary 44
2.11 Problems 45
References 47
Appendix 2A: Vectors 47
2A.1 Products 48
2A.2 Derivatives 48
Appendix 2B: Electromagnetic Units 50

3 Reflection and Refraction 52

3.1 Introduction 52
3.2 Reflection and Transmission at a Discontinuity 53

3.3 Laws of Reflection and Refraction 56
3.4 Fresnel's Formula 58
3.4.1 σ Case (Perpendicular Polarization) 60
3.4.2 π Case (Parallel Polarization) 61
3.5 Reflected and Transmitted Energy 63
3.6 Normal Incidence 65
3.7 Polarization by Reflection 68
3.8 Total Reflection 69
3.9 Reflection from a Conductor 73
3.10 Summary 75
3.11 Problems 77
References 78

4 **Interference** 79

4.1 Introduction 79
4.2 Addition of Waves 79
4.2.1 Trigonometric Approach 80
4.2.2 Complex Approach 81
4.2.3 Vector Approach 81
4.3 Interference 82
4.4 Young's Interference 86
4.5 Dielectric Layer 89
4.5.1 Fizeau Fringes 91
4.5.2 Color Fringes 91
4.5.3 Haidinger's Fringes 94
4.5.4 Antireflection coating 94
4.5.5 Newton's Rings 95
4.6 Michelson Interferometer 97
4.7 Interference by Multiple Reflection 101
4.7.1 Fabry–Perot Interferometer 106
4.8 Summary 112
4.9 Problems 114
References 116
Appendix 4A: Multilayer Dielectric Coatings 116
4A.1 Vector Approach 117
4A.2 Matrix Approach 121

5 **Geometrical Optics** 126

5.1 Introduction 126
5.2 Eikonal Equation 128
5.3 Fermat's Principle 129
5.4 Applications of Fermat's Principle 132
5.4.1 Law of Reflection 132
5.4.2 Law of Refraction 133
5.4.3 Propagation through an Optical System 134
5.5 Lens Design and Matrix Algebra 135
5.6 Geometrical Optics of Resonators 141

5.7 Guided Waves 146
5.7.1 End Coupling 149
5.7.2 Guided Modes 153
5.7.3 Propagation Vector Formalism 157
5.7.4 Solution for Asymmetric Guide 158
5.7.5 Solution for Symmetric Guide 160
5.7.6 Cutoff Condition 162
5.7.7 Coupling into Guided Wave Modes 162
5.7.7.1 Fiber Coupling 163
5.7.7.2 Evanescent Wave Coupling 164
5.8 Lagrangian Formulation of Optics 165
5.8.1 Hamilton's Principle 166
5.8.2 Rectilinear Propagation 168
5.8.3 Law of Refraction 168
5.9 Propagation in a Graded-Index Optical Fiber 169
5.10 Summary 177
5.11 Problems 179
References 182
Appendix 5A: The *ABCD* Matrix 183
5A.1 Thin-Lens Equation 184
5A.2 Optical Invariant 184
5A.3 Lensmaker's Equation 185
5A.4 Gaussian Formalism 186
5A.5 Newtonian Formalism 186
5A.6 Principal Planes 187
5A.6.1 Nodal Points 189
5A.7 Aperture Stop and Pupil 189
Appendix 5B: Aberrations 191
5B.1 Wavefront Aberration Coefficients 194
5B.1.1 Optical Path Difference 196
5B.1.2 Transverse Ray Coefficients 197
5B.2 Spherical Aberrations 198
5B.2.1 Ray Intercept Plot 201
5B.3 Coma 202
5B.3.1 Optical Sine Theorem 204
5B.3.2 Spot Diagram 205
5B.4 Astigmatism 206
5B.5 Field Curvature 207
5B.6 Distortion 209
5B.7 Aberration Reduction 210
5B.7.1 Coddington Shape Factor 210
5B.7.2 Coddington Position Factor 211

6 Fourier Analysis 214

6.1 Introduction 214
6.2 Fourier Series 215
6.2.1 DC Term 216
6.2.2 Cosine Series 216
6.2.3 Sine Series 217

6.2.4 Even and Odd Functions 217
6.2.5 Exponential Representation 218
6.3 Periodic Square Wave 219
6.4 The Fourier Integral 222
6.4.1 Dirichlet Conditions 223
6.4.2 Evaluation of the Fourier Transform 224
6.5 Rectangular Pulse 225
6.6 Pulse Modulation Wave Trains 227
6.7 Dirac Delta Function 230
6.8 Replication and Sampling 234
6.9 Correlation 237
6.10 Convolution Integrals 239
6.11 Linear System Theory 242
6.12 Fourier Transforms in Two Dimensions 244
6.13 Summary 246
6.14 Problems 249
References 251
Appendix 6A: Fourier Transform Properties 252
6A.1 Linearity 252
6A.2 Scaling 252
6A.3 Shifting 252
6A.4 Conjugation 252
6A.5 Differentiation 252
6A.6 Convolution 252
6A.7 Parseval's Theorem 252
6A.8 Correlation 252
6A.9 Common Fourier Transform Pairs 253
6A.10 Convolution Properties 253

7 **Dispersion** 254

7.1 Introduction 254
7.2 Stiff Strings 256
7.3 Group Velocity 258
7.4 Dispersion of Guided Waves 264
7.5 Material Dispersion 266
7.5.1 Conductive Gas 267
7.5.1.1 Plasma Frequency 268
7.5.2 Molecular Gas 271
7.5.3 Dense Dielectric 277
7.5.4 Metals 283
7.6 Plasmons and Polaritons 283
7.7 Lorenz–Lorentz Law 287
7.8 Signal Velocity, Superluminal Propagation 289
7.9 Summary 293
7.10 Problems 295
References 298
Appendix 7A: Chromatic Aberrations 298

8 Coherence 306

8.1 Introduction 306
8.2 Photoelectric Mixing 307
8.3 Interference Spectroscopy 310
8.4 Fourier Transform Spectroscopy 312
8.4.1 Gaussian Spectral Distribution 315
8.5 Fringe Contrast and Coherence 316
8.6 Temporal Coherence Time 318
8.7 Autocorrelation Function 320
8.8 Spatial Coherence 323
8.8.1 A Line Source 325
8.8.2 van Cittert–Zernike Theorem 327
8.9 Spatial Coherence Length 329
8.10 Stellar Interferometer 330
8.11 Intensity Interferometry 332
8.12 Summary 335
8.13 Problems 338
References 340

9 Diffraction and Gaussian Beams 341

9.1 Introduction 341
9.2 Huygens' Principle 343
9.2.1 Rectilinear Propagation 345
9.2.2 Law of Reflection 345
9.2.3 Snell's Law 346
9.3 Fresnel Formulation 347
9.4 The Obliquity Factor 350
9.4.1 Approximate Solutions of the Huygens–Fresnel integral 355
9.5 Gaussian Beams 356
9.6 Higher-Order Gaussian Modes and Bessel Beams 363
9.6.1 Hermite–Gaussian Waves 363
9.6.2 Laguerre–Gaussian Waves 364
9.6.3 Bessel Beams 364
9.7 Beam Propagation 365
9.7.1 The *ABCD* Law 366
9.7.2 Thin Lens 368
9.7.3 Fabry–Perot Resonator 369
9.7.4 Laser Cavity 371
9.8 Summary 372
9.9 Problems 373
References 375
Appendix 9A: Fresnel–Kirchhoff Diffraction 375
Appendix 9B: Rayleigh–Sommerfeld Formula 381

10 Fraunhofer Diffraction 384

10.1 Introduction 384
10.2 Fraunhofer Diffraction 385

10.3 Fourier Transforms via a Lens 388
10.4 Plane Wave Representation 391
10.5 Diffraction by a Rectangular Aperture 392
10.6 Diffraction by a Circular Aperture 394
10.7 Array Theorem 398
10.8 *N* Rectangular Slits 401
10.8.1 Young's Double Slit 402
10.8.2 The Diffraction Grating 403
10.9 Summary 411
10.10 Problems 413
References 416
Appendix 10A: Abbe Theory and Optical Processing 417
10A.1 Introduction 417
10A.2 Abbe's Theory of Imaging 417
10A.3 Amplitude Spatial Filtering 428
10A.4 Apodization 432
10A.5 Phase Filtering 434
10A.6 Phase and Amplitude Filter 437
Appendix 10B: Imaging 442
10B.1 Introduction 442
10B.2 Incoherent Imaging 443
10B.2.1 Resolution Criteria 444
10B.2.2 Optical Transfer Function 447
10B.2.3 Modulation Transfer Function 450
10B.3 Coherent Imaging 451
10B.4 Computational Imaging 455

11 Fresnel Diffraction 459

11.1 Introduction 459
11.2 Fresnel Approximation 460
11.3 Rectangular Apertures 464
11.4 Fresnel Zones 469
11.4.1 Incident Plane Wave 476
11.5 Circular Aperture 477
11.5.1 Intensity near the Aperture 478
11.5.2 Off-Axis Intensity 479
11.6 Opaque Screen 479
11.7 Zone Plate 481
11.8 Pinhole Camera 484
11.9 Fermat's Principle 485
11.10 Comparison of Techniques 486
11.11 Summary 487
11.12 Problems 488
References 489
Appendix 11A: Babinet's Principle 490
11A.1 Fraunhofer Diffraction 490
11A.2 Fresnel Diffraction 491

Appendix 11B: Fresnel Integral Solutions 493
11B.1 Table of Fresnel Integrals 493
11B.2 Cornu Spiral 495

12 Periodic Index of Refraction 496

12.1 Introduction 496
12.2 Holography 497
12.3 Holographic Recording 502
12.4 Off-Axis Holography 505
12.4.1 Recording 505
12.4.2 Development 507
12.4.3 Reconstruction 507
12.5 Spatial Spectrum of Off-Axis Holograms 508
12.6 Classification of Holograms 511
12.7 Diffraction Efficiency 512
12.8 Holography and Zone Plates 513
12.9 Resolution Requirements 517
12.10 Coherence Requirements 519
12.10.1 Temporal Coherence 519
12.10.2 Spatial Coherence 520
12.11 Photonic Crystals 522
12.11.1 Maxwell's Equations for Sinusoidal ε 522
12.12 The Bloch (or Floquet) Theorem 524
12.13 Multilayer Photonic Crystal 526
12.13.1 Photonic Crystal Defect 527
12.14 Numerical Techniques 529
12.15 Two-Dimensional Periodic Structure 530
12.16 Three-Dimensional Periodic Structure 533
12.17 Fabrication Defects 533
12.18 Natural Photonic Crystals 533
12.19 Summary 533
12.20 Problems 536
References 539
Appendix 12A: Phase Holograms 540
Appendix 12B: VanderLugt Filter 543

13 Anisotropy 549

13.1 Introduction 549
13.2 Dichroic Polarizers 551
13.2.1 Crystals 551
13.2.2 Wire Grids 552
13.2.3 Polaroid Sheet 552
13.3 Reflection Polarizers 554
13.3.1 Brewster's Angle Polarizers 555
13.3.2 Interference Polarizers 555
13.4 Polarization by Birefringence 559
13.5 Optical Indicatrix 564

13.6 Fresnel's Equation 568
13.6.1 Transverse Waves 569
13.6.2 Interdependence of **D** and **H** 569
13.6.3 Fresnel's Equation 569
13.7 Retarder 572
13.7.1 Quarter-Wave Plate 572
13.7.2 Compensator 573
13.7.3 Rhomb 575
13.8 Mueller Calculus 575
13.9 Jones Calculus 577
13.10 Optical Activity 578
13.11 Summary 585
13.12 Problems 586
References 588
Appendix 13A: Tensors 588
13A.1 Scalars 589
13A.2 Vectors 589
13A.3 Second-Rank Tensors 589
13A.4 Higher-Rank Tensors 589
13A.5 Coordinate Transformations 590
13A.6 Geometrical Representation 590
13A.7 Crystal Symmetry 591
Appendix 13B: Poynting Vector in an Anisotropic Dielectric 592
13B.1 Ray Ellipsoid 594
Appendix 13C: Normal Surfaces 595
13C.1 Biaxial Crystal 595
13C.2 Uniaxial Crystal 595
13C.3 Refraction in Crystals 597
Appendix 13D: Ray Surfaces 599

14 Optical Modulation 602

14.1 Introduction 602
14.2 Electro-optic Effect 604
14.3 Electro-optic Indicatrix 606
14.3.1 Pockels Effect 606
14.3.2 Kerr Effect 610
14.4 Amplitude Modulation 612
14.4.1 Kerr Modulation 615
14.4.2 Pockels Modulation 616
14.5 Modulator Design 621
14.5.1 Kerr Modulator 621
14.5.2 Pockels Modulator 621
14.5.3 Longitudinal Modulator 622
14.5.4 Transverse Modulator 623
14.6 Magneto-optic Effect 623
14.6.1 Cotton–Mouton and Voigt Effects 629
14.7 Photoelastic Effect 630

14.8 Acousto-optics 634
14.8.1 Bragg Scattering 637
14.8.2 Raman–Nath Scattering 638
14.8.3 Acousto-optic Modulator 640
14.8.4 Spectrum Analyzer 643
14.8.5 Acousto-optic Beam Deflector 644
14.9 Summary 646
14.10 Problems 648
References 649
Appendix 14A: Pockels and Kerr Tensors 649
Appendix 14B: Phenomenological Acousto-optic Theory 653
14B.1 Bragg Region 656
14B.2 Raman–Nath Region 659
Appendix 14C: Acoustic Figure of Merit 662

15 Nonlinear Optics 664

15.1 Introduction 664
15.2 Nonlinear Polarization 667
15.3 Nonlinear Optical Coefficient 670
15.4 Symmetry Properties 671
15.5 Wave Propagation in a Nonlinear Medium 673
15.6 Conservation of Energy 676
15.6.1 Manley–Rowe Relation 676
15.7 Conservation of Momentum 677
15.7.1 Poynting Vector 677
15.7.2 Phase Matching 680
15.8 Second-Harmonic Generation 682
15.9 Methods of Phase Matching 684
15.9.1 Quasi-Phase Matching (QPM) 684
15.9.2 Total Internal Reflection 685
15.9.3 Dielectric Waveguide 686
15.9.4 Noncollinear Phase Matching 686
15.9.5 Birefringent Phase Matching 688
15.10 Phase Conjugation 693
15.11 Summary 698
15.12 Problems 701
References 702
Appendix 15A: Nonlinear Optical Medium 703
15A.1 Introduction 703
15A.2 Generalized Linear Theory 704
15A.3 Nonlinear Equation of Motion 704
15A.4 Perturbation Technique 705
15A.5 Second-Order Nonlinearity 708
Appendix 15B: Miller's Rule 710
Appendix 15C: Nonlinear Polarization in the 32 Point Group 711

Index 714

Wave Theory

1

1.1 Introduction 1
1.2 Traveling Waves 1
1.3 Wave Equation 4
1.4 Transmission of Energy 5
1.5 Three Dimensions 6
1.6 Attenuation of Waves 10
1.7 Summary 11
1.8 Problems 12

1.1 Introduction

The theory of wave motion is an important mathematical model in many areas of physics. A large number of seemingly unrelated phenomena can be explained using the solution of the wave equation, the basic equation of wave theory. The equation was first used to describe the behavior of a vibrating violin string by Daniel Bernoulli and Jean d'Alembert in the eighteenth century. Wave theory is a fundamental part of modern quantum theory, and solutions of the wave equation are used to explain a number of classical phenomena. Familiarity with wave theory developed in the study of light will aid in the understanding of such diverse physical processes as water waves, vibrating drums and strings, traffic dynamics, and seismic waves.

Mathematically, the basis of wave theory is a second-order partial differential equation called the wave equation. In this chapter, a traveling wave on a string will be used to find the functional form of a one-dimensional wave. Following a discussion of the energy associated with the traveling wave, the one-dimensional model associated with the string illustration will be expanded to three dimensions. The displacement of the wave discussed in this chapter is assumed to be a scalar function, and the theory is called a *scalar wave theory*. In Chapter 2, the vector wave theory will be discussed.

Christian Huygens (**1629–1695**) developed the wave theory of light in 1678. **Isaac Newton** (**1642–1727**) proposed a counter theory based on a particle view of light. Because of Newton's scientific stature, only a few scientists during the eighteenth century, for example **Leonard Euler** (**1707–1783**) and **Benjamin Franklin** (**1706–1790**), accepted the wave theory and rejected Newton's particle theory. **Thomas Young** (**1773–1829**) in 1801 and **Augustin Jean Fresnel** (**1788–1827**) in 1814 utilized experiments to demonstrate interference and diffraction of light and presented a theoretical explanation of these experiments in terms of the wave theory. Fresnel was able to explain rectilinear propagation using the wave theory, thereby removing Newton's main objection to the theory. Acceptance of Fresnel's theory came very slowly, and the final rejection of Newton's theory did not come until the measurement of the speed of light in water and air by **Jean Bernard Léon Foucault** (**1819–1868**). The velocity measurements were a key element in the rejection of Newton's theory, because the particle theory required the speed of light in a medium to exceed the speed of light in a vacuum in order to explain refraction. The measurements by Foucault showed the propagation velocity in a vacuum to exceed that in water.

1.2 Traveling Waves

Before the equation of motion of a wave is discussed, a mathematical expression for a wave will be obtained. We will assume that a disturbance propagates without change along a string and that each point on the string undergoes simple harmonic motion. This assumption will allow us to obtain a simple mathematical expression for a wave that will be used to define the parameters that characterize a wave.

Modern Optics. Second Edition. B. D. Guenther.

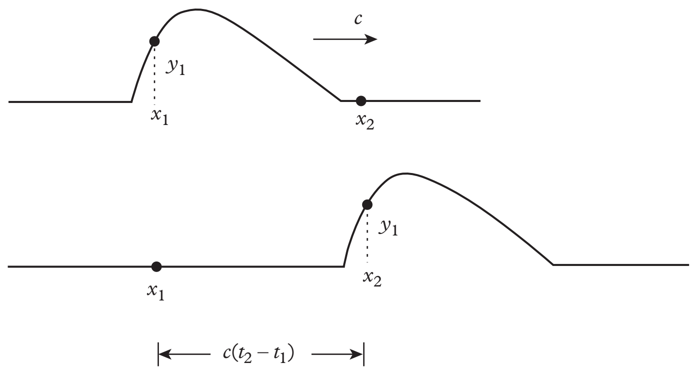

Figure 1.1 *Propagation of a pulse on a guitar string. We assume that the amplitude does not change as the pulse propagates along the string.*

A guitar string is plucked, creating a pulse, which travels to the right and left along the x-axis at a constant speed c. In Figure 1.1, the pulse traveling toward the right is shown. The pulse's amplitude is defined as

$$y \equiv f(x, t)$$

and equals y_1 at position x_1 and time t_1. This amplitude travels a distance $c(t_2 - t_1)$ to the right of x_1 and is shown graphically in Figure 1.1.

Assume the pulse does not change in amplitude as it propagates:

$$f(x_1, t_1) = f(x_2, t_2),$$

where $x_2 = x_1 + c(t_2 - t_1)$. If the function has the form

$$y = f(ct - x), \tag{1.1}$$

then the requirement that the pulse does not change is satisfied because

$$\begin{aligned} f(x_1, t_1) &= f(ct_1 - x_1), \\ f(x_2, t_2) &= f(ct_2 - x_2) = f(ct_2 - x_1 c(t_2 - t_1)) \\ &= f(ct_1 - x_1). \end{aligned}$$

Using the same reasoning, we can show that an unchanging pulse traveling from right to left, along the x-axis, with speed c is described by

$$y = g(x + ct).$$

The expression $y = f(ct - x)$ is a shorthand notation to denote a function that contains x and t only in the combination $ct - x$; i.e., the function can contain combinations of the form $2(ct - x)$, $t \pm x/c$, $x - ct$, $(ct - x)^2$, $\sin(ct - x)$, etc. but not expressions such as $2ct - x$ or $ct^2 - x^2$.

To the assumption of an unchanging propagating disturbance is now added the requirement that each point on the guitar string oscillate transversely, i.e., perpendicular to the direction of propagation, with simple harmonic motion:

$$m\frac{d^2y}{dt^2} + sy = 0,$$

where s is often called the spring constant. The string in Figure 1.1 lies along the x-axis, and the harmonic motion will be in the y-direction. The point on the string at the origin ($x = 0$) undergoes simple harmonic motion with amplitude Y and frequency ω. Note that throughout this book, we shall use the angular frequency $\omega = 2\pi\nu$, where the linear frequency ν is defined as the reciprocal of the period of oscillation:

$$\nu = \frac{1}{T} = \frac{1}{2\pi}\sqrt{\frac{s}{m}}.$$

The equation describing the motion of the origin is

$$y = Y\cos\omega t.$$

The origin acts as a source of a continuous train of pulses (a wave train) moving to the right.

A function of $ct - x$ that will reduce to harmonic motion at $x = 0$ is

$$y = f(ct - x) = Y\cos\left[\frac{\omega}{c}(ct - x)\right].$$

This is called a ***harmonic wave***.

A number of different notations are used for a harmonic wave; the one used in this book involves a constant

$$k = \frac{\omega}{c}, \tag{1.2}$$

called the ***propagation constant*** or the ***wavenumber***. The harmonic wave is written

$$y = Y\cos(\omega t - kx). \tag{1.3}$$

The change in the value of x for which the phase $\omega t - kx$ changes by 2π and thus leaves the right-hand side of (1.3) unchanged is the ***spatial period*** and is called the ***wavelength*** λ. Let $x_2 = x_1 + \lambda$, so that

$$\omega t - kx_2 = \omega t - kx_1 - k\lambda = \omega t - kx_1 - 2\pi;$$

thus,

$$k = \frac{2\pi}{\lambda}. \tag{1.4}$$

Since $k = \omega/c = 2\pi\nu/c$, we also have the relationship $c = \nu\lambda$.

To determine the speed of the wave in space, a point on the wave is selected, and the time it takes to go some distance is measured. This is equivalent to asking how fast a given value of phase propagates in space. Assume that in the time $\Delta t = t_2 - t_1$, the disturbance y_1 travels a distance $\Delta x = x_2 - x_1$, as shown in Figure 1.1. Since the disturbance at the two points is the same, i.e., y_1, the phases must be equal:

$$\omega t - kx = \omega(t + \Delta t) - k(x + \Delta x),$$

$$\frac{\Delta x}{\Delta t} = \frac{\omega}{k}.$$

In the limit as $\Delta t \to 0$, we obtain the ***phase velocity***

$$c \equiv \frac{dx}{dt} = \frac{\omega}{k}.$$

The adjective "phase" is used because this velocity describes the motion of a preselected phase of the wave. Another method that can be used to obtain the propagation speed associated with a wave is to define the phase velocity using the following result from partial differential calculus:

$$\left(\frac{\partial x}{\partial t}\right)_y = -\frac{\left(\frac{\partial y}{\partial t}\right)_x}{\left(\frac{\partial y}{\partial x}\right)_t} = \frac{\omega}{k}. \tag{1.5}$$

This equation may be verified by applying it to (1.3).

1.3 Wave Equation

We may write the wave equation as

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 y}{\partial t^2}. \tag{1.6}$$

This differential equation describes a number of physical situations: a vibrating string, sound waves, a vibrating drum head, elastic waves in solids such as seismic waves in the earth, electric signals in a cable, and electromagnetic waves—the subject of this book.

We said earlier that the motion of a vibrating string was described by equations of the form (1.1):

$$y_1 = f(ct - x) \quad \text{or} \quad y_2 = g(ct + x).$$

To show that y_1 is a solution, we substitute it into the wave equation (we will denote derivatives with respect to the argument of the function f, i.e., $ct - x$, by a prime):

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial (ct-x)}\frac{\partial (ct-x)}{\partial t} = cf', \ \frac{\partial f}{\partial x} = \frac{\partial f}{\partial (ct-x)}\frac{\partial (ct-x)}{\partial x} = (-1)f',$$

$$\frac{\partial^2 f}{\partial t^2} = c^2 f'', \ \frac{\partial^2 f}{\partial x^2} = f'',$$

$$\frac{\partial^2 f}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = f'' - \frac{1}{c^2}\left(c^2 f''\right) = 0.$$

The same procedure is used to prove that $g(ct + x)$ is a solution.

Another important property of the wave equation is the ***superposition principle***. This states that the sum of two solutions is also a solution. Proof of the superposition principle is the objective of Problem 1.10.

1.4 Transmission of Energy

Each element of a string, with a harmonic wave propagating along it, moves up and down in the y-direction, undergoing simple harmonic motion. This can be seen by selecting a coordinate position to observe, say x_1, and replacing kx_1 in (1.3) by a constant δ, i.e., $y = Y\cos(\omega t - \delta)$. Equation (1.3) now assumes the same form as the equation for a harmonic oscillator:

$$x = A\cos(\omega_0 t + \delta), \tag{1.7}$$

$$v = -\omega_0 A\ \sin(\omega_0 t + \delta), \tag{1.8}$$

$$a = -\omega_0^2 A\ \cos(\omega_0 t + \delta). \tag{1.9}$$

The elements of the string do not move in the direction of wave motion (the x-direction). Even though the elements are not translated along the direction of propagation, energy is transmitted. An aid to understanding how this occurs can be obtained by imagining yourself at the end of a long line of people waiting to purchase tickets. To communicate with a friend at the front of the line, you could pass a note from hand to hand until it reaches your friend at the front of the line. No one need move in the direction of your friend, and yet the note arrives.

Energy is transmitted by a wave in the same way as the note. To discover the wave characteristics that determine the energy transmitted by a wave, consider a string, under tension T, as shown in Figure 1.2: point "a" on the string has work done on it by the point to its left and it does work on the point on its right. The work done on point "a" is

$$dW = F\,dy = (T\sin\theta)(v\,dt).$$

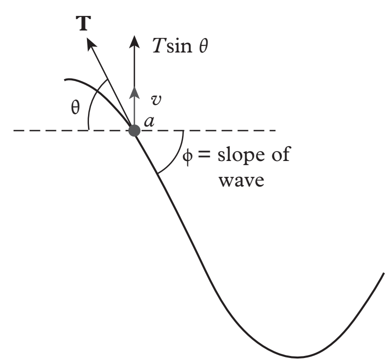

Figure 1.2 *Forces acting on point "a" of a string.*

The work done is equal to the change in energy, allowing us to write the instantaneous power transmitted as

$$P = \frac{dE}{dt} = Tv\ \sin\theta.$$

From (1.3), we can write

$$\frac{dE}{dt} = YT\sin\theta[-\omega\sin(\omega t - kx)].$$

We continue to assume small-amplitude waves, so that the tension at point "a" in Figure 1.2 is parallel to the slope of the wave at "a":

$$\sin\theta \approx \tan\theta \approx -\tan\phi = -\frac{dy}{dx} = -kY\sin(\omega t - kx),$$

$$\frac{dE}{dt} = T[-kY\sin(\omega t - kx)][-\omega Y\sin(\omega t - kx)],$$

$$\frac{dE}{dt} = Tk\omega Y^2\sin^2(\omega t - kx).$$

The average power transmitted in one period ($2\pi/\omega$) is

$$P_{\text{avg}} = \langle E\rangle = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} P\,dt = \frac{Tk\omega^2 Y^2}{2\pi}\int_0^{2\pi/\omega}\sin^2(\omega t - kx)\ dt,$$

$$P_{\text{avg}} = \frac{Tk\omega Y^2}{2} = \frac{T\omega^2 Y^2}{2c}. \tag{1.10}$$

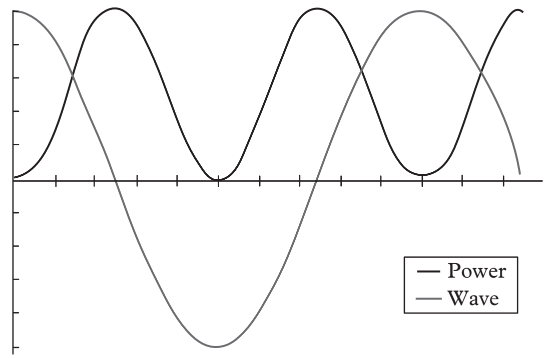

Figure 1.3 *A traveling wave and its instantaneous power are plotted with their amplitudes normalized to one.*

For a given string, the tension T and the phase velocity c are constants, and the average power transmitted is proportional to the square of the frequency and the square of the amplitude. In Figure 1.3, we plot the wave and its instantaneous power, demonstrating that the power travels along with the wave.

1.5 Three Dimensions

While the wave traveling along a string can be described with only one dimension, we will need a three-dimensional model in our discussion of light. The one-dimensional model (1.6) generalizes to a three-dimensional wave equation

$$\frac{\partial^2 f(\mathbf{r},t)}{\partial x^2} + \frac{\partial^2 f(\mathbf{r},t)}{\partial y^2} + \frac{\partial^2 f(\mathbf{r},t)}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 f(\mathbf{r},t)}{\partial t^2}, \tag{1.11}$$

or, in vector notation [see (2A.8) in Appendix 2A],

$$\nabla^2 f(\mathbf{r},t) = \frac{1}{c^2}\frac{\partial^2 f(\mathbf{r},t)}{\partial t^2}. \tag{1.12}$$

A generalized harmonic wave solution of (1.12) is

$$f(\mathbf{r},t) = E(\mathbf{r})\cos[\omega t - \mathcal{S}(\mathbf{r})].$$

At a fixed time, the surfaces for which $\mathcal{S}(\mathbf{r}) = \text{constant}$ are called ***wavefronts***. If $E(\mathbf{r})$, the amplitude of the wave, is a constant over the wavefront, then the wave is said to be ***homogeneous***; if not, then the wave is ***inhomogeneous***. The average power defined by (1.10) is defined for the three-dimensional case as the average power transmitted per unit cross-sectional area. This quantity is called the ***intensity*** of the wave.

Another notation we will find useful in the study of waves is complex notation. The generalized solution of the wave equation can be expressed in complex notation as

$$f(\mathbf{r},t) = \mathcal{E}(\mathbf{r})e^{i\omega t},$$

where

$$\mathcal{E}(\mathbf{r}) = E(\mathbf{r})e^{-i[\mathcal{S}(\mathbf{r})+\delta]}$$

and δ is an arbitrary phase angle. If we substitute $f(\mathbf{r}, t)$ into the wave equation, we obtain

$$\frac{\partial^2 f(\mathbf{r}, t)}{\partial t^2} = -\omega^2 \mathcal{E}(\mathbf{r}) e^{i\omega t},$$
$$\nabla^2 f(\mathbf{r}, t) = e^{i\omega t} \nabla^2 \mathcal{E}(\mathbf{r}).$$

Using the relationship $\omega/c = k$, the wave equation becomes

$$(\nabla^2 + k^2)\mathcal{E}(\mathbf{r}) = 0. \tag{1.13}$$

If we are interested in the spatial properties of the wave but not the temporal properties, we need only seek solutions of this equation, which is called the ***Helmholtz equation***.

The evaluation of the phase velocity of the three-dimensional wave is more complicated than it was for the one-dimensional case because we must monitor the motion of a surface. As we show below, it is possible to derive the phase velocity of an arbitrary wavefront. In this book, however, we will restrict our consideration to two simple wavefronts: a plane (called a plane wave) and a sphere (called a spherical wave).

To calculate the phase velocity of an arbitrary wavefront, we must follow a point of constant phase as the wave moves through space and time. If the phase at $(\mathbf{r}, t)$ is equal to the phase at $(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t)$, then

$$\omega t - \mathcal{S}(\mathbf{r}) = \omega t + \omega\, \Delta t - \mathcal{S}(\mathbf{r} + \Delta\mathbf{r}),$$
$$\omega \Delta t - \left[\frac{\mathcal{S}(\mathbf{r} + \Delta\mathbf{r}) - \mathcal{S}(\mathbf{r})}{\Delta\mathbf{r}} \right] \Delta\mathbf{r} = 0,$$

and in the limit as $\Delta t \Delta\mathbf{r} \to 0$, the term in square brackets becomes the gradient of $\mathcal{S}(\mathbf{r})$:

$$\omega dt - \nabla\mathcal{S}(\mathbf{r}) \cdot d\mathbf{r} = 0. \tag{1.14}$$

If we define a unit vector $\hat{\mathbf{n}}$ in the direction of $d\mathbf{r}$, so that $d\mathbf{r} = \hat{\mathbf{n}}\, ds$, where ds is the distance between the surfaces $\mathcal{S}(\mathbf{r})$ and $\mathcal{S}(\mathbf{r} + d\mathbf{r})$, then the velocity of the surface is

$$\frac{ds}{dt} = \frac{\omega}{\hat{\mathbf{n}} \cdot \nabla\mathcal{S}(\mathbf{r})}.$$

Figure 1.4 shows these parameters for a plane surface. This derivative has its smallest value when $\hat{\mathbf{n}}$ is normal to the wavefront,

$$\hat{\mathbf{n}} = \frac{\nabla\mathcal{S}(\mathbf{r})}{|\nabla\mathcal{S}(\mathbf{r})|}.$$

We then have an equation for the phase velocity of the wavefront, c:

$$\frac{ds}{dt} = c = \frac{\omega}{|\nabla\mathcal{S}(\mathbf{r})|}. \tag{1.15}$$

Note that the phase velocity is not a vector. We must exercise care in assigning any physical significance to this velocity.

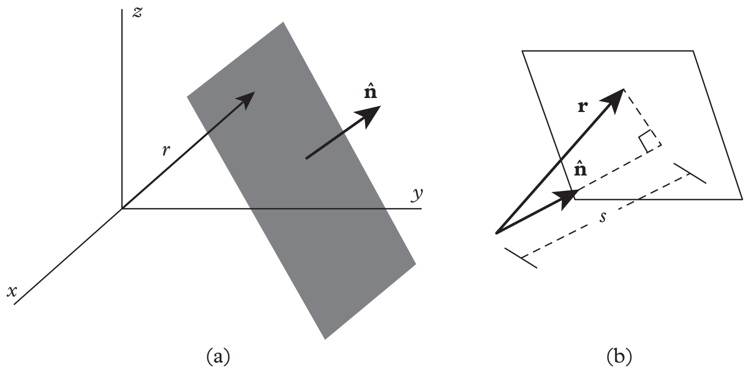

Figure 1.4 *(a) A plane wave. Its normal is the unit vector* $\hat{\mathbf{n}}$*, which points in the direction of propagation. A surface of constant phase is the shaded plane passing through the point defined by the vector* $\mathbf{r}$*. (b) The projection of* $\mathbf{r}$ *on the plane's normal defines the distance s from the origin.*

Mathematically, a plane is described by the equation

$$\mathbf{r} \cdot \hat{\mathbf{n}} = s$$

where $\mathbf{r}$ is the position vector of a point in the plane, $\hat{\mathbf{n}}$ is a unit vector normal to the plane, and s is a constant equal to the distance from the origin to the plane, as shown in Figure 1.4. For a wavefront to be a plane, we must have $S(\mathbf{r}) = k(\mathbf{n} \cdot \mathbf{r})$. For convenience, we define the wavevector $\mathbf{k} = k\hat{\mathbf{n}}$, and we write our plane wave solution of the wave equation as

$$f(\omega t - \mathbf{k} \cdot \mathbf{r}) + g(\omega t + \mathbf{k} \cdot \mathbf{r}).$$

The harmonic plane wave solution is

$$f(\mathbf{r}, t) = E(\mathbf{r}) \cos(\omega t - \mathbf{k} \cdot \mathbf{r}). \tag{1.16}$$

In complex notation, the harmonic plane wave is written as

$$f(\mathbf{r}, t) = E(\mathbf{r}) e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}. \tag{1.17}$$

We will find this harmonic plane wave important because, as we will see later, any three-dimensional wave can be written as a combination of plane waves of different amplitudes, directions, and frequencies.

The second three-dimensional wave we will find useful is one with spherical symmetry:

$$f(\mathbf{r}, t) = f(r, \theta, \phi, t) = f(r, t),$$

where

$$r = \sqrt{x^2 + y^2 + z^2}.$$

For example, a point source located at the origin would produce a wave with a wavefront that is a sphere. In this case, the wavefront is given by $S(\mathbf{r}) = S(r) = kr =$ constant (the equation for a sphere). The wave equation can be obtained by converting from rectangular to spherical coordinates (we only have to obtain the r component, because f is not a function of θ and ϕ if it has spherical symmetry):

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} = \frac{x}{r}\frac{\partial f}{\partial r},$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial r}\left(\frac{x}{r}\frac{\partial f}{\partial r}\right)\frac{\partial r}{\partial x} = \frac{x}{r}\left[\frac{x}{r}\frac{\partial^2 f}{\partial r^2} + \frac{\partial f}{\partial r}\frac{\partial}{\partial r}\left(\frac{x}{r}\right)\right],$$

$$= \frac{x^2}{r^2}\frac{\partial^2 f}{\partial r^2} + \left(\frac{1}{r} - \frac{x^2}{r^3}\right)\frac{\partial f}{\partial r}. \tag{1.18}$$

We do the same for the derivatives with respect to y and z and add the results to (1.18):

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \frac{1}{r^2}\frac{\partial^2 f}{\partial r^2}\left(x^2 + y^2 + z^2\right) + \frac{3}{r}\frac{\partial f}{\partial r} - \frac{x^2 + y^2 + z^2}{r^3}\frac{\partial f}{\partial r},$$

$$= \frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r}.$$

The wave equation for spherically symmetric solutions becomes

$$\frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r} = \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2},$$

or, equivalently,

$$\frac{\partial^2}{\partial r^2}(rf) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}(rf).$$

This is the one-dimensional wave equation with a general solution

$$rf(r,t) = f(ct - r) + g(ct + r).$$

The harmonic spherical wave is

$$f(r,t) = \frac{A}{r}\cos(\omega t - kr). \tag{1.19}$$

It is easy to find physical examples of the spherical wave. One realization that everyone has seen is the water wave formed on the surface of a still pool by a drop of water striking the surface, as shown in Figure 1.5. As the wave moves out from its source, it forms circles of ever increasing radii. This water wave can be thought of as a representation of the intersection of the spherical wavefront and a plane that cuts through the wave and contains the source. If we assume that there are no losses, then the total power in the wave is a constant, P_0, given by the product of the power per unit area, P, and the area of the spherical wave, $4\pi r^2$:

$$P_0 = P\left(4\pi r^2\right).$$

This leads to the conclusion that the power per unit area is inversely proportional to the radius r of the spherical wave's phase front:

$$P = \frac{P_0}{4\pi r^2}.$$

Figure 1.5 *Spherical wavefront. Uploaded from Flickr by Jacopo Werther, taken by Davide Restivo.* http://commons.wikimedia.org/wiki/File:Water_drop_impact_on_a_water-surface_-_(5).jpg.

Since r is also the distance that the wave has traveled from the source, we have that the power per unit area in a spherical wave is inversely proportional to the square of the distance the wave has propagated. The power per unit area of a spherical wave can also be obtained by using (1.19) in (1.10) and is given by

$$P \propto \frac{A^2}{r^2}.$$

Conservation of energy and the wave model both yield the inverse square law behavior.

1.6 Attenuation of Waves

Physical systems dissipate a wave's energy as it propagates. We can take the effect of a dissipative force into account by adding a loss or damping term to the wave equation. We will use the same functional form for the loss term as is used for a damped harmonic oscillator:

$$m\frac{d^2y}{dt^2} = -sy - b\frac{dy}{dt},$$

where b is a positive constant of proportionality, with units of kg/s, called the ***damping constant***. The solutions of this equation behave as shown in Figure 1.6. The new equation of motion is then

$$\frac{\partial^2 f}{\partial x^2} - \gamma\frac{\partial f}{\partial t} - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = 0, \tag{1.20}$$

where $\gamma = b/m$ is the resistance per unit time. Is this equation still a wave equation? We can find out by testing if the one-dimensional representation of a plane wave (1.17),

$$f(x,t) = Ae^{i(\omega t - kx + \phi)}, \tag{1.21}$$

is a solution of (1.20). If we differentiate (1.21), we obtain

$$\frac{\partial^n f}{\partial x^n} = (-ik)^n Ae^{i(\omega t - kx + \phi)} = (-ik)^n f(x,t), \tag{1.22}$$

$$\frac{\partial^n f}{\partial t^n} = (i\omega)^n Ae^{i(\omega t - kx + \phi)} = (i\omega)^n f(x,t). \tag{1.23}$$

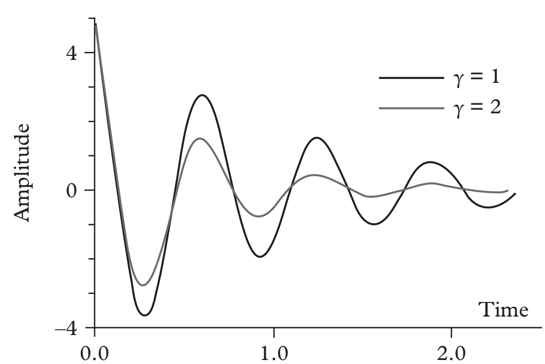

Figure 1.6 *Damped harmonic oscillator with two different size loss terms.*

Here the usefulness of complex notation becomes apparent. The untidy equations obtained when differentiating sine and cosine functions are avoided. Substituting the above results into (1.20) yields

$$k^2 = \left(\frac{\omega}{c}\right)^2 - i\omega\gamma.$$

This type of equation is called a ***dispersion equation*** or ***dispersion relation***. We see that for (1.20) to have a plane wave solution of the form (1.21), k must be complex. If we write the complex k as

$$\tilde{k} = \kappa_1 - i\kappa_2 \tag{1.24}$$

and substitute this into (1.21), we obtain

$$f(x,t) = Ae^{-\kappa_2 x} e^{i(\omega t - \kappa_1 x + \phi)}.$$

A solution of (1.21) is a harmonic wave whose amplitude is attenuated as it propagates in the positive x-direction. κ_2 is an attenuation constant whose reciprocal equals the distance the wave will propagate before its amplitude falls to $1/e$ of the value it had at $x = 0$.

1.7 Summary

We developed a one-dimensional model of a wave whose properties are described by the wave equation

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 f(x,t)}{\partial t^2}.$$

We extended this equation to three dimensions:

$$\nabla^2 f(\mathbf{r},t) = \frac{1}{c^2}\frac{\partial^2 f(\mathbf{r},t)}{\partial t^2}.$$

A special case of the wave equation called the Helmholtz equation

$$(\nabla^2 + k^2)\mathcal{E}(\mathbf{r}) = 0$$

was introduced for those situations where only the spatial properties of the wave are to be discussed.

The most important solution of the wave equation for the discussions in this book is the harmonic plane wave

$$f(\mathbf{r},t) = E(\mathbf{r})\cos(\omega t - \mathbf{k}\cdot\mathbf{r}),$$

or, in complex notation,

$$f(\mathbf{r},t) = E(\mathbf{r})e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})}.$$

The solution of the wave equation is modified if the propagation medium has some loss mechanism. A plane harmonic wave propagating in the x-direction then has an amplitude that is attenuated as it propagates:

$$f(x,t) = Ae^{-\kappa_2 x} e^{i(\omega t - \kappa_1 x + \phi)}$$

where

$$\tilde{k} = \boldsymbol{\kappa}_1 - i\boldsymbol{\kappa}_2$$

is the complex wavevector for the wave propagating in the dissipative medium.

The average power per unit area (equivalently the energy per unit time per unit area) was defined as the intensity of the wave and was shown to be proportional to the amplitude squared.

1.8 Problems

1.1. Assume a sine wave with an amplitude of 10 cm and a wavelength of 200 cm moving with a velocity of 100 cm/s:
   (a) What is the frequency, both ν and ω?
   (b) What is the value of k?
   (c) What is the wave equation?
   (d) Assume the left end is at the origin and moving down at $t = 0$. What is the equation of motion at the left end?
   (e) What is the equation of motion of the point 150 cm to the right of the origin?

1.2. The wavelength of light ranges from 390 to 780 nm and its velocity (in a vacuum) is about 3×10^8 m/s. What is the corresponding frequency range?

1.3. Given the complex number

$$\tilde{z}_1 = \frac{1}{2i}(1 - 4i),$$

what are its real part and its modulus? Calculate

$$\tilde{z}_1 \tilde{z}_1^*.$$

What is the imaginary part of

$$\tilde{z}_2 = \frac{e^{i\omega t} - e^{-i\omega t}}{2} \quad ?$$

1.4. If

$$f(x,t) = Ae^{i(kx - \omega t)},$$

what are the expressions for $\mathcal{Re}\{|f|^2\}$ and $[\mathcal{Re}\{f\}]^2$?

1.5. Show that

$$s(x,t) = Ae^{-(2x+3t)^2}$$

is a solution of the wave equation. What is the velocity? In what direction is the wave moving?

1.6. The thickness of a human hair is about 4×10^{-2} mm. Compare its dimension with that of a light wave.

1.7. Find the direction of travel of the following two waves:

$$s(x,t) = A\sin(kx - \omega t),$$
$$s(x,t) = A\cos(\omega t - kx).$$

1.8. If we write the wavefunction in complex notation

$$s(x,t) = Ae^{i\phi},$$

show that s is unchanged when its phase increases or decreases by 2π.

1.9. Given the wave

$$\varphi(x,t) = 10 \ \cos\left\{2\pi\left[\frac{x}{2 \times 10^{-7}} - (1.5 \times 10^{15})t\right]\right\},$$

determine its speed, wavelength, and frequency in MKS units.

1.10. If s_1 and s_2 are solutions of the wave equation

$$\frac{\partial^2 s}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 s}{\partial t^2},$$

prove the superposition principle, i.e., that $as_1 + bs_2$ is also a solution, where a and b are both constants.

1.11. A simple harmonic oscillator has a mass m of 0.01 kg and a force constant s of 36 N/m. At $t = 0$, the mass is displaced 50 mm to the right of its equilibrium position and is moving to the right at a speed of 1.7 m/s. Calculate (a) the frequency, (b) the period, (c) the amplitude, and (d) the phase constant.

1.12. The HCl molecule, to first order, acts like a simple harmonic oscillator. The hydrogen, which does all the vibrating, has a mass of 1.67×10^{-27} kg, and the force constant is 840 N/m. What is the vibration frequency?

1.13. Assume the motion of an oscillator is described by

$$y = 2\cos 6\pi t + \sin 5\pi t.$$

If the mass of the oscillator is 10 g, what are the maximum and minimum kinetic energies?

1.14. When playing a guitar, harmonics of the open string are generated by placing a finger lightly on the string above the 12th fret (this is at the halfway point between the supports of the guitar string, i.e., the nut and the saddle). What other positions, in terms of string length, would be useful for harmonic generation?

1.15. A mass of 10 g is attached to a spring and causes it to stretch 1 cm downward. If the attached mass is moved downward another 1 cm and released, what will be its period and maximum velocity?

1.16. A spring has a spring constant $k = 2.5$ N/m and a damping force constant $b = 1$ kg/m. A mass of 25 g is attached to the spring, and the spring is released at $t = 0$ from a position at $x = -5$ cm. Find (a) the frequency and (b) the epoch angle (the initial phase at $t = 0$). (c) Write an expression for the displacement as a function of time of the oscillator.

1.17. Find the times when the maximum displacements occur and the values of the maximum displacement for the first three excursions of the oscillator in Problem 1.16.

1.18. A particle of mass m that is constrained to move along the x-axis experiences a force toward the origin equal to mkx. Upon release, the particle undergoes simple harmonic motion about the origin with a period of 6 s. At time $t = 2$ s, the particle reaches the origin. At $t = 3$ s, the particle is found to have a velocity of 5 cm/s. (a) What is the force constant? (b) What is the maximum distance from the origin reached by the particle?

1.19. A string 3 m long and weighing 300 g is held under a tension of 10 N. How long does it take a wave to travel the string's length?

1.20. A string is held under tension by passing it over a pulley and attaching a 2 kg weight. An 8 m-long segment of the string, which weighs 600 g, has a wave of amplitude 10 cm and wavelength 3 m traveling along its length. (a) What is the velocity of the wave? (b) What is the maximum transverse velocity of a point on the wave?

1.21. A string with mass per unit length $\mu = 0.1$ kg/m and $T = 90$ N has a wave of amplitude 30 cm and frequency 1 Hz propagating along it. How much energy is transmitted by the wave?

Electromagnetic Theory

2

2.1	**Introduction**	15
2.2	**Maxwell's Equations**	16
2.3	**Free Space**	19
2.4	**Wave Equation**	19
2.5	**Transverse Waves**	21
2.6	**Interdependence of E and B**	22
2.7	**Energy Density and Flow**	24
2.8	**Polarization**	28
2.9	**Propagation in a Conducting Medium**	40
2.10	**Summary**	44
2.11	**Problems**	45
	References	47
	Appendix 2A: Vectors	47
	Appendix 2B: Electromagnetic Units	50

2.1 Introduction

The scalar wave theory that we discussed in Chapter 1 was applied to the study of light before the development of the theory of electromagnetism. At that time, it was assumed that light waves were longitudinal, in analogy with sound waves; i.e., the wave displacements were in the direction of propagation. A further assumption, that light propagated through some type of medium, was made because the scientists of that time approached all problems from a mechanistic point of view. The scalar theory was successful in explaining diffraction (see Chapter 9), but problems arose in interpretation of the effects of polarization in interference experiments (discussed in Chapter 4). Young was able to resolve the difficulties by suggesting that light waves could be transverse, like the waves on a vibrating string. Using this idea, Fresnel developed a mechanistic description of light that could explain the amount of reflected and transmitted light from the interface between two media (see Chapter 3).

Independent of this activity, the theory of electromagnetism was under development. **Michael Faraday (1791–1867)** observed in 1845 that a magnetic field would rotate the plane of polarization of light waves passing through the magnetized region. This observation led Faraday to associate light with electromagnetic radiation, but he was unable to quantify this association. Faraday attempted to develop electromagnetic theory by treating the field as lines pointing in the direction of the force that the field would exert on a test charge. The lines were given a mechanical interpretation with a tension along each line and a pressure normal to the line. **James Clerk Maxwell (1831–1879)** furnished a mathematical framework for Faraday's model in a paper read in 1864 and published a year later [1]. In this paper, Maxwell identified light as "an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws" and demonstrated that the propagation velocity of light was given by the electromagnetic properties of the material.

Maxwell was not the first to recognize the connection between the electromagnetic properties of materials and the speed of light. Kirchhoff (**Gustav Robert Kirchhoff: 1824–1887**) recognized in 1857 that the speed of light could be obtained from electromagnetic properties. Riemann (**Georg Friedrich Bernhard Riemann: 1826–1866**), in 1858, assumed that electromagnetic forces propagated at a finite velocity and derived a propagation velocity given by the electromagnetic properties of the medium. However, it was Maxwell who demonstrated that the electric and magnetic fields are waves that travel at the speed of light. It was not until 1887 that an experimental observation of electromagnetic waves, other than light, was obtained by **Heinrich Rudolf Hertz (1857–1894)**.

The classical electromagnetic theory is successful in explaining all of the experimental observations to be discussed in this book. There are, however, experiments that cannot be explained by classical wave theory, especially those conducted at short wavelengths or very low light levels. Quantum electrodynamics (QED) is capable of predicting the outcome of all optical experiments; its shortcoming is that it does not explain why or how. An excellent elementary introduction to QED has been written by Richard Feynman [2].

Modern Optics. Second Edition. B. D. Guenther.

In this chapter, we will borrow, from electromagnetic theory, Maxwell's equations and Poynting's theorem to derive properties of light waves. Details of the origins of these fundamental electromagnetic relationships are not needed for our study of light but can be obtained by consulting any electricity and magnetism text (e.g., [3]).

The basic properties that will be derived are

- the wave nature of light;
- the fact that light is a transverse wave;
- the velocity of light in terms of fundamental electromagnetic properties of materials;
- the relative magnitude of the electric and magnetic fields and relationships between the two fields;
- the energy associated with a light wave.

The concept of polarized light and a geometrical construction used to visualize its behavior will be introduced. In the seventeenth century, Hooke postulated that light waves might be transverse, but his idea was forgotten. Young and Fresnel made the same postulate in the nineteenth century and accompanied this with a theoretical description of light based on transverse waves. Forty years later, Maxwell proved that light must be a transverse wave and that **E** and **H**, for a plane wave in an isotropic medium with no free charges and no currents, are mutually perpendicular and lie in a plane normal to the direction of propagation, **k**.

Convention requires that we use the electric vector to label the direction of the electromagnetic wave's polarization. The direction of the displacement vector is called the *direction of polarization*, and the plane containing the direction of polarization and the propagation vector is called the *plane of polarization*. The selection of the electric field is not completely arbitrary—except for relativistic situations, when $v \approx c$, the interaction of the electromagnetic wave with matter will be dominated by the electric field. Both a vector and a matrix notation for describing polarization will be presented in this chapter, but details on the manipulation of light polarization will not be discussed until Chapter 14. The chapter will conclude with a discussion of the propagation of light in a conducting medium.

2.2 Maxwell's Equations

The bases of electromagnetic theory are Maxwell's equations. They allow the derivation of the properties of light. In our study of optics, we will treat these equations as axioms, but we provide the reader with a reference source here that can be consulted if information on the origin of the equations is desired.

In rationalized MKS units, Maxwell's equations are as follows.

2.2.1 Gauss's Law

2.2.1.1 Gauss's (Coulomb's) Law for the Electric Field

Coulomb's law provides a means for calculating the force between two charges (see Chapter 2 of [3]),

$$\mathbf{F} = \frac{q_0}{4\pi\varepsilon_0}\int \frac{dq}{r^2}\hat{\mathbf{n}},$$

where dq is the charge on an infinitesimal surface and $\hat{\mathbf{n}}$ is a unit vector in the direction of the line connecting the charges q_0 and dq. The electric field

$$\mathbf{E} = \frac{\mathbf{F}}{q_0}$$

is obtained using Coulomb's law (see Chapter 3 of [3]). We view this field as Michael Faraday did, as lines of flux, called lines of force, originating on positive charges and terminating on negative charges. Gauss's Law states that the quantity of charge contained within a closed surface is equal to the number of flux lines passing outward through the surface (see Chapter 4 of [3]). This view of the electric field leads to

$$\nabla \cdot \mathbf{D} = \rho, \tag{2.1}$$

where ρ is the charge density and $\mathbf{D}$ is the electric displacement (see Chapter 10, Section 5 of [3]). The use of the displacement allows the equation to be applied to any material.

2.2.1.2 Gauss's Law for the Magnetic Field

Charges at rest led to (2.1). Charges in motion, i.e., a current $\mathbf{i}$ or a current density $\mathbf{J}$, create a magnetic field $\mathbf{B}$ (see Chapter 14 of [3]). As we did for the electric field, we treat the magnetic field as flux lines, called lines of induction, and we assume that the current density is a constant so that $\nabla \cdot \mathbf{J} = 0$. This leads to (see Chapter 16 of [3])

$$\nabla \cdot \mathbf{B} = 0. \tag{2.2}$$

The zero is due to the fact that the magnetic equivalent of a single charge has never been observed.

2.2.2 Faraday's Law

The previous two equations are associated with electric and magnetic fields that are constant with respect to time. The next equation, an experimentally derived equation, deals with a magnetic field that is time-varying or equivalently a conductor moving through a static magnetic field:

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0. \tag{2.3}$$

In terms of the concept of flux, it states that an electric field around a circuit is associated with a change in the magnetic flux contained within the circuit.

2.2.3 Ampére's Law (Law of Biot and Savart)

An electric charge in motion creates a magnetic field around its path. The law of Biot and Savart allows us to calculate the magnetic field at a point located a distance $\mathbf{R}$ from a conductor carrying a current density $\mathbf{J}$. Ampère's law is the inverse relationship used to calculate the current in a conductor due to the magnetic field contained in a loop about the conductor. Neither relationship is adequate when the current is a function of time. Maxwell's major contribution to physics was to observe that the addition of a *displacement current* to Ampère's law allowed fluctuating currents to be explained. The relationship became (see Chapter 21 of [3])

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \tag{2.4}$$

As discussed in Appendix 2B, the constants in Maxwell's equations depend on the units used. Many optics books use c.g.s. units, which result in a form for Maxwell's equations shown in Appendix 2B.

2.2.4 Constitutive Relations

The dynamic responses of atoms and molecules in the propagation medium are taken into account through what are called the *constitutive relations*:

$$\mathbf{D} = f(\mathbf{E}),$$
$$\mathbf{J} = g(\mathbf{E}),$$
$$\mathbf{B} = h(\mathbf{H}).$$

Here we will assume that the functional relations are independent of space and time, and we will write the constitutive relations as

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad \varepsilon = \text{dielectric constant},$$
$$\mathbf{J} = \sigma\mathbf{E}, \quad \sigma = \text{conductivity (Ohm's law)},$$
$$\mathbf{B} = \mu\mathbf{H}, \quad \mu = \text{permeability},$$

where the constants ε, σ, and μ contain the description of the material. Later, we will explore the effects resulting from the constitutive relations having a temporal or a spatial dependence.

Often, **D** and **B** are defined as

$$\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}, \tag{2.5}$$
$$\mathbf{B} = \mu_0\mathbf{H} + \mathbf{M},$$

where **P** is the polarization and **M** is the magnetization. This formulation emphasizes that the internal field of a material is due not only to the applied field but also due to a field created by the atoms and molecules that make up the material. We will find (2.5) useful in Chapters 7 and 15. We will not use the relationship involving **M** in this book.

By manipulating Maxwell's equations, we can obtain a number of the properties of light, such as its wave nature, the fact that it is a transverse wave, and the relationship between the **E** and **B** fields. We will make a number of simplifying assumptions about the medium in which the light is propagating to allow a quick derivation of the properties of light. Later, we will see what happens if we modify these assumptions.

2.3 Free Space

We assume that the light is propagating in a medium that we will call *free space* and that is

(1) uniform: ε and μ have the same value at all points;
(2) isotropic: ε and μ do not depend upon the direction of propagation;
(3) nonconducting: $\sigma = 0$, and thus $\mathbf{J} = 0$;
(4) free from charge: $\rho = 0$;
(5) nondispersive: ε and μ are not functions of frequency, i.e., they have no time dependence.

Our definition departs somewhat from other definitions of free space in that we include in the definition not only the vacuum, where $\varepsilon = \varepsilon_0$ and $\mu = \mu_0$, but also dielectrics, where $\sigma = 0$ but the other electromagnetic constants can have arbitrary values.

Using the above assumptions, Maxwell's equations and the constitutive relations simplify to

$$\nabla \cdot \mathbf{E} = 0, \qquad (2.6a) \qquad\qquad \nabla \cdot \mathbf{B} = 0, \qquad (2.6b)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \qquad (2.6c) \qquad\qquad \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \qquad (2.6d)$$

$$\mathbf{B} = \mu \mathbf{H}, \qquad (2.6e) \qquad\qquad \varepsilon \mathbf{E} = \mathbf{D}. \qquad (2.6f)$$

These simplified equations can now be used to derive some of the basic properties of a light wave.

2.4 Wave Equation

To find how the electromagnetic wave described by (2.6) propagates in free space, Maxwell's equations must be rearranged to display explicitly the time and coordinate dependence. Using (2.6e, f), we can rewrite (2.6d) as

$$\frac{1}{\mu}\nabla \times \mathbf{B} = \varepsilon \frac{\partial \mathbf{E}}{\partial t}.$$

The curl of (2.6c) is taken and the magnetic field dependence is eliminated by using the rewritten (2.6d):

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right) = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = -\frac{\partial}{\partial t}\left(\varepsilon\mu \frac{\partial \mathbf{E}}{\partial t}\right).$$

The assumption that ε and μ are independent of time allows this equation to be rewritten as

$$\nabla \times (\nabla \times \mathbf{E}) = -\varepsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

Using the vector identity (2A.12) from Appendix 2A, we can write

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\varepsilon\mu \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

Because free space is free of charge, $\nabla \cdot \mathbf{E} = 0$, giving us

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{2.7}$$

We can use the same procedure to obtain

$$\nabla^2 \mathbf{B} = \mu\varepsilon \frac{\partial^2 \mathbf{B}}{\partial t^2}. \tag{2.8}$$

These equations are wave equations, with the wave's velocity being given by

$$v = \frac{1}{\sqrt{\mu\varepsilon}}. \tag{2.9}$$

The connection of the velocity of light with the electric and magnetic properties of a material was one of the most important results of Maxwell's theory. In a vacuum,

$$\mu_0\varepsilon_0 = \left(4\pi \times 10^{-7}\right)\left(8.8542 \times 10^{-12}\right) = 1.113 \times 10^{-17}\ \mathrm{s^2/m^2},$$

$$\frac{1}{\sqrt{\mu_0\varepsilon_0}} = 2.998 \times 10^{8}\ \mathrm{m/s} = c. \tag{2.10}$$

In a material, the velocity of light is less than c. We can characterize a material by defining the *index of refraction*, the ratio of the speed of light in a vacuum to its speed in a medium:

$$n = \frac{c}{v} = \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}. \tag{2.11}$$

The data in Table 2.1 demonstrate that if magnetic materials are not considered, then $\mu/\mu_0 \approx 1$, so that

$$n = \sqrt{\frac{\varepsilon}{\varepsilon_0}}\ .$$

Table 2.1 *Representative magnetic permeabilities*

Material	μ/μ_0	Class
Silver	0.99998	Diamagnetic
Copper	0.99999	Diamagnetic
Water	0.99999	Diamagnetic
Air	1.00000036	Paramagnetic
Aluminum	1.000021	Paramagnetic
Iron	5000	Ferromagnetic
Nickel	600	Ferromagnetic

Table 2.2 *Selected indices of refraction*

Material	n (yellow light)	$\sqrt{\varepsilon/\varepsilon_0}$ (static)
Air	1.000294	1.000295
CO_2	1.000449	1.000473
C_6H_6 (benzene)	1.482	1.489
He	1.000036	1.000034
H_2	1.000131	1.000132

The data displayed in Table 2.2 demonstrate that, at least for some materials, the theory agrees with experimental results. The materials whose indices are listed in Table 2.2 have been specially selected to demonstrate good agreement; we will see in Chapter 7 that the assumption that ε, μ, and σ are independent of frequency results in a theory that neglects the response time of the system to the electromagnetic signal.

2.5 Transverse Waves

Hooke postulated, in the seventeenth century, that light waves might be transverse, but his idea was forgotten. In the nineteenth century, Young and Fresnel made the same postulate and provided a theoretical description of light based on transverse waves. Forty years later, Maxwell proved that light must be a transverse wave. We can demonstrate the transverse nature of light by substitution of the plane wave solution of the wave equation into Gauss's law:

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0.$$

To complete the demonstration, we consider the divergence of the electric component of the plane wave. We will examine only the x-coordinate of the divergence in detail:

$$\begin{aligned}\frac{\partial E_x}{\partial x} &= \frac{\partial}{\partial x}\left(E_{0x}e^{i(\omega t-\mathbf{k}\cdot\mathbf{r}+\phi)}\right) = iE_{0x}e^{i(\omega t-\mathbf{k}\cdot\mathbf{r}+\phi)}\frac{\partial}{\partial x}(\omega t - k_x x - k_y y - k_z z + \phi)\\ &= -ik_x E_x.\end{aligned}$$

We easily obtain similar results for E_y and E_z, allowing the divergence of $\mathbf{E}$ to be rewritten as a dot product of $\mathbf{k}$ and $\mathbf{E}$. Gauss's law for the electric field states that the divergence of $\mathbf{E}$ is zero, which for a plane wave can be written

$$\nabla \cdot \mathbf{E} = -i\mathbf{k} \cdot \mathbf{E} = 0. \tag{2.12}$$

If the dot product of two vectors $\mathbf{E}$ and $\mathbf{k}$, is zero, then the vectors $\mathbf{E}$ and $\mathbf{k}$ must be perpendicular [see (2A.1) in Appendix 2A]. In the same manner, substituting the plane wave into $\nabla \cdot \mathbf{B} = 0$ yields $\mathbf{k} \cdot \mathbf{B} = 0$. Therefore, Maxwell's equations require light to be a transverse wave; i.e., the vector displacements $\mathbf{E}$ and $\mathbf{B}$ are perpendicular to the direction of propagation, $\mathbf{k}$.

2.6 Interdependence of E and B

The electric and magnetic fields are not independent, as we can see by continuing our examination of the plane wave solutions of Maxwell's equations. First, let us calculate several derivatives of the plane wave. We will need

$$\begin{aligned}\frac{\partial \mathbf{B}}{\partial t} &= \frac{\partial}{\partial t}\mathbf{B}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r}+\phi)} = i\mathbf{B}\frac{\partial}{\partial t}(\omega t - \mathbf{k}\cdot\mathbf{r}+\phi)\\ &= i\omega\mathbf{B}\end{aligned} \tag{2.13}$$

and, similarly,

$$\frac{\partial \mathbf{E}}{\partial t} = i\omega\mathbf{E}. \tag{2.14}$$

A simple expression for the curl of **E**, $\nabla \times \mathbf{E}$, can be obtained when we use the derivatives just calculated. The expression for the curl of **E** is given by (2A.7) from Appendix 2A and is rewritten here:

$$\nabla \times \mathbf{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\hat{\mathbf{i}} + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\hat{\mathbf{j}} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\hat{\mathbf{k}}.$$

The terms making up the x-component of the curl are

$$\frac{\partial E_z}{\partial y} = E_{0z}\frac{\partial}{\partial y}e^{i(\omega t - \mathbf{k}\cdot\mathbf{r}+\phi)} = -ik_y E_z$$

and

$$\frac{\partial E_y}{\partial z} = -ik_z E_y.$$

By evaluating each component, we find that the curl of **E** for a plane wave is

$$\nabla \times \mathbf{E} = -i\mathbf{k} \times \mathbf{E}. \tag{2.15}$$

A similar derivation leads to the curl of **B** for a plane wave:

$$\nabla \times \mathbf{B} = -i\mathbf{k} \times \mathbf{B}. \tag{2.16}$$

With these vector operations on a plane wave defined, we can evaluate (2.6c) for a plane wave. The left-hand side of

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

is replaced with (2.15) and the right-hand side by (2.13), resulting in an equation connecting the electric and magnetic fields:

$$-i\mathbf{k} \times \mathbf{E} = -i\omega\mathbf{B}.$$

Using the relationship between ω and $\mathbf{k}$ given by (1.2) from Chapter 1 and the relationship for the wave velocity in terms of the electromagnetic properties of the material, (2.9), we can write

$$\frac{\sqrt{\mu\varepsilon}}{k}\mathbf{k} \times \mathbf{E} = \mathbf{B}. \tag{2.17}$$

A second relationship between the magnetic and electric fields can be generated by using the same procedure to rewrite

$$\nabla \times \mathbf{B} = \mu\varepsilon\frac{\partial \mathbf{E}}{\partial t}$$

for a plane wave as

$$-i\mathbf{k} \times \mathbf{B} = i\varepsilon\mu\omega\mathbf{E};$$

that is,

$$\frac{1}{k\sqrt{\mu\varepsilon}}\mathbf{k} \times \mathbf{B} = -\mathbf{E}. \tag{2.18}$$

From the definition of the cross product given by (2A.2) in Appendix 2A, we see that the electric and magnetic fields are perpendicular to each other, are in phase, and form a right-handed coordinate system with propagation direction **k** (see Figure 2.1).

If we are only interested in the magnitude of the two fields, we can use (2.11) to write

$$n|\mathbf{E}| = c|\mathbf{B}|. \tag{2.19}$$

In a vacuum, we take $n = 1$ in (2.19). For our plane wave, the ratio of the field magnitudes is

$$\frac{|\mathbf{E}|}{|\mathbf{H}|} = \sqrt{\frac{\mu}{\varepsilon}}. \tag{2.20}$$

This ratio has units of ohms ($\mu \Rightarrow \mathrm{ml/Q^2}$, $\varepsilon \Rightarrow \mathrm{Q^2t^2/ml^3}$, and $\Omega \Rightarrow \mathrm{ml^2/Q^2t}$) and is called the impedance of the medium. In a vacuum,

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 377\ \Omega.$$

When the ratio is a real quantity, as it is here, **E** and **H** are in phase.

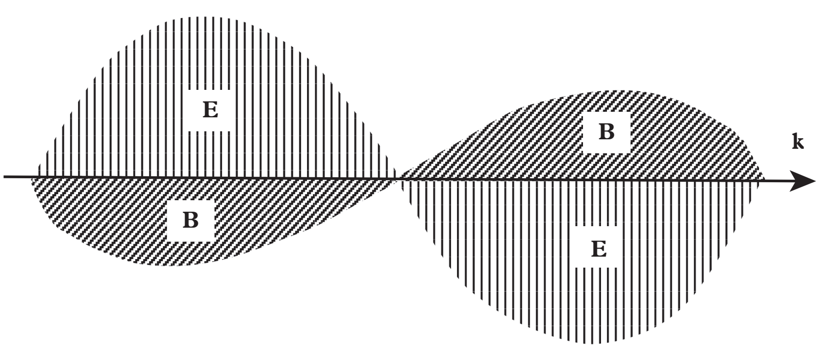

Figure 2.1 *Graphical representation of an electromagnetic plane wave. Note that* **E** *and* **B** *are perpendicular to each other and individually perpendicular to the propagation vector* **k**, *are in phase, and form a right-handed coordinate system as required by (2.17) and (2.18).*

2.7 Energy Density and Flow

We saw in our discussion of waves propagating along strings that the power transmitted by a wave is proportional to the square of the amplitude of the wave. Any text on electromagnetic theory (see, e.g., Chapter 21 of [3]) demonstrates that the energy density (in J/m^3) associated with an electromagnetic wave is given by

$$U = \frac{\mathbf{D} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{H}}{2}. \tag{2.21}$$

We can simplify (2.21) by using the simple constitutive relations $\mathbf{D} = \varepsilon \mathbf{E}$ and $\mathbf{B} = \mu \mathbf{H}$, if they apply to the propagation medium:

$$U = \frac{1}{2}\left(\varepsilon E^2 + \frac{B^2}{\mu}\right) = \frac{1}{2}\left(\varepsilon + \frac{1}{\mu c^2}\right) E^2.$$

In a vacuum, further simplification is possible:

$$U = \varepsilon_0 E^2 = \frac{B^2}{\mu_0}.$$

John Henry Poynting (1852–1914), an English professor of physics at Mason Science College, now the University of Birmingham, demonstrated that the presence of both an electric and a magnetic field at the same point in space results in a flow of the field energy. This fact is called the Poynting theorem, and the flow is completely described by the *Poynting vector*

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \tag{2.22}$$

The Poynting vector has units of J/(m^2·s). We will use a plane wave to determine some of the properties of this vector. Since $\mathbf{S}$ will involve terms quadratic in $\mathbf{E}$, it will be necessary to use the real form of $\mathbf{E}$ (see Problem 1.4). We have

$$\mathbf{H} = \frac{\mathbf{B}}{\mu} = \frac{\sqrt{\mu\varepsilon}}{\mu k} \mathbf{k} \times \mathbf{E},$$

where

$$\mathbf{E} = \mathbf{E}_0 \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi).$$

Then,

$$\begin{aligned} \mathbf{S} &= \frac{\sqrt{\mu\varepsilon}}{\mu k} \mathbf{E}_0 \times (\mathbf{k} \times \mathbf{E}_0) \cos^2(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi) \\ &= \frac{n}{\mu c} |\mathbf{E}_0|^2 \frac{\mathbf{k}}{k} \cos^2(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi). \end{aligned} \tag{2.23}$$

Note that the energy is flowing in the direction of propagation (indicated by the unit vector $\mathbf{k}/k$).

We normally do not detect **S** at the very high frequencies associated with light ($\approx 10^{15}$ Hz) but rather detect a temporal average of **S** taken over a time T determined by the response time of the detector used. We must obtain the time average of **S** to relate theory to actual measurements. The time average of **S** is called the *flux density* and has units of W/m^2. We will call this quantity the ***intensity*** of the light wave,

$$I = |\langle \mathbf{S} \rangle| = \left| \frac{1}{T} \int_{t_0}^{t_0+T} \mathbf{A} \cos^2(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi)\, dt \right|, \tag{2.24}$$

where we have defined

$$\mathbf{A} = \frac{n}{\mu c} |\mathbf{E}_0|^2 \frac{\mathbf{k}}{k}$$

to simplify the notation.

The units used for the flux density are a confusing mess in optics. One area of optics is interested in measuring the physical effects of light, and the measurement of energy is called *radiometry*. In radiometry, the flux density is called the *irradiance*, with units of W/m^2. Another area of optics is interested in the psychophysical effects of light, and the measurement of energy is called *photometry*. For this group, the flux density is called *illuminance*, with units of a lumen/m^2 (lm/m^2) or a lux. Each of these two group has its own set of units, but both desire to measure the energy flow of a field that is not well defined in frequency or phase. Much research in modern optics belong to a third area that is associated with the use of a light source that has both a well-defined frequency and a well-defined phase—the laser. In this area of optics, common usage defines the time-averaged flux density as the intensity. In this book, all of the waves discussed are uniquely defined in terms of the electric field and the electromagnetic properties of the material in which the wave is propagating. To emphasize that the results of our theory are immediately applicable only to a light source with a well-defined frequency and phase, we will use the term *intensity* for the magnitude of the Poynting vector.

We will assume that **k** is independent of time over the period T:

$$\langle \mathbf{S} \rangle = \frac{\mathbf{A}}{\omega T} \int_{t_0\omega}^{(t_0+T)\omega} \cos^2(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi)\, d(\omega t).$$

Using the trigonometric identity

$$\cos^2\theta = \frac{1}{2}(1 + \cos 2\theta)$$

and evaluating the integral results in the expression

$$\langle \mathbf{S} \rangle = \frac{\mathbf{A}}{2} + \frac{\mathbf{A}}{4\omega T}\left[\sin 2(\omega t_0 + \omega T - \mathbf{k} \cdot \mathbf{r} + \phi) - \sin 2(\omega t_0 - \mathbf{k} \cdot \mathbf{r} + \phi)\right]. \tag{2.25}$$

The largest value that the term in square brackets can assume is 2. The period T is the response time of the detector to the light wave. Normally, it is much longer than the period of light oscillations, so $\omega T \gg 1$ and we can neglect the second term in (2.25). As an example, suppose our detection system has a 1 GHz bandwidth, yielding a response time of $T = 10^{-9}$ s (the reciprocal of the bandwidth). Green light has a frequency of $\nu = 6 \times 10^{14}$ Hz or $\omega \approx 4 \times 10^{15}$ rad/s. With these values, $\omega T = 4 \times 10^5$, and the neglected term would be no larger than 10^{-6} of the first term. Therefore, in optics, the assumption that $\omega T \gg 1$ is reasonable and allows the average Poynting vector to be written as

$$\langle \mathbf{S} \rangle = \frac{\mathbf{A}}{2} = \frac{n}{2\mu c} |\mathbf{E}_0|^2 \frac{\mathbf{k}}{k}. \tag{2.26}$$

The energy per unit time per unit area depends on the square of the amplitude of the wave.

The energy calculation was done with plane waves of **E** and **H** that are in phase. We will later see that materials with nonzero conductivity $\sigma \neq 0$ will yield a complex impedance because **E** and **H** are no longer in phase. If the two waves are 90° out of phase, then the integral in (2.24) will contain $\sin x \ \cos x$ as its integrand, resulting in $\langle \mathbf{S} \rangle = 0$. Therefore, no energy is transmitted.

In quantum mechanics, the energy of light is carried by discrete particles called photons. If the light has a frequency ν, then the energy of a photon is $h\nu$. The intensity of the light is equal to the number of photons, striking unit area in unit time, N, multiplied by the energy of an individual photon:

$$I = Nh\nu.$$

The intensity of a 10 mW HeNe laser beam, 2 mm in diameter, is

$$I = \frac{\text{power}}{\text{area}} = \frac{10^{-2}}{\pi \left(10^{-3}\right)^2} = 3.18 \times 10^3 \ \mathrm{W\,m^{-2}}.$$

The number of photons in this beam can be calculated once we know that the wavelength of the light is 632.8 nm:

$$N = \frac{I}{h\nu} \cdot \text{area} = \frac{10^{-2}\lambda}{hc} = \frac{10^{-2}\left(632.8 \times 10^{-9}\right)}{\left(6.6 \times 10^{-34}\right)\left(3 \times 10^8\right)} = 32 \times 10^{15}.$$

We can get a perspective on how large this number is by comparing it with the number of molecules in a mole of a molecule, i.e., Avogadro's constant

$$N_A = 6.02214129 \times 10^{23}.$$

For carbon-12, a mole is 12 g.

The energy crossing a unit area A in a time Δt is contained in a volume $A(v\Delta t)$ (in a vacuum $v = c$), as shown in Figure 2.2. To find the magnitude of this energy, we must multiply this volume by the average energy density $|\langle \mathbf{S} \rangle|$. Thus, we expect the energy flow to be given by

$$|\langle \mathbf{S} \rangle| = \frac{\text{energy}}{A\Delta t} \propto \frac{Av\Delta t \langle U \rangle}{A\Delta t} = v \langle U \rangle.$$

We may use the definitions of the wave velocity

$$v = \frac{1}{\sqrt{\mu\varepsilon}}$$

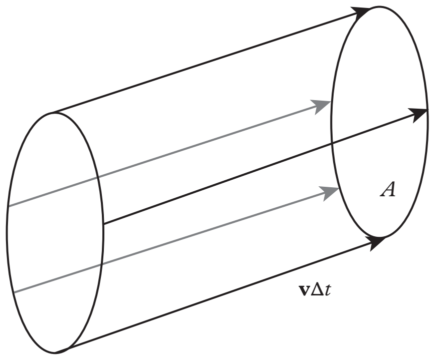

Figure 2.2 *The energy of a wave crossing a unit area A in a time* Δt.

and index of refraction $n = c/v$ to rewrite (2.26) as

$$|\langle \mathbf{S} \rangle| = \frac{\varepsilon v E_0^2}{2} = v \langle U \rangle, \tag{2.27}$$

giving the expected result that the energy is flowing through space at the speed of light in the medium. The relationship defined by (2.27), (energy flow) = (wave velocity) · (energy density), is a general property of waves.

At the Earth's surface, the flux density of full sunlight is 1.34×10^3 J/(m^2·s). It is not completely correct to do so, but if we associate this flux with the time average of the Poynting vector, then the electric field associated with the sunlight is $E_0 = 10^3$ V/m.

Our discussion of the time average of the Poynting vector provides an opportunity to discover one of the advantages of the use of complex notation. To obtain the time average of the product of two waves A and B, where

$$A = \mathcal{Re}\{\tilde{A}\} = \mathcal{Re}\{A_0 e^{i(\omega t + \phi_1)}\},$$
$$B = \mathcal{Re}\{\tilde{B}\} = \mathcal{Re}\{B_0 e^{i(\omega t + \phi_2)}\},$$

we use

$$\mathcal{Re}\{\tilde{z}\} = x = r\cos\phi = \frac{\tilde{z} + \tilde{z}^*}{2},$$
$$\mathcal{Im}\{\tilde{z}\} = y = r\sin\phi = \frac{\tilde{z} - \tilde{z}^*}{2i}$$

to write the average over one period as

$$\langle AB \rangle = \frac{1}{T}\int_0^T \left(\frac{\tilde{A} + \tilde{A}^*}{2}\right)\left(\frac{\tilde{B} + \tilde{B}^*}{2}\right) dt,$$
$$(\tilde{A} + \tilde{A}^*)(\tilde{B} + \tilde{B}^*) = \tilde{A}\tilde{B} + \tilde{A}^*\tilde{B}^* + \tilde{A}\tilde{B}^* + \tilde{A}^*\tilde{B},$$

where

$$\tilde{A}\tilde{B} = A_0 B_0 e^{i(2\omega t+\phi_1+\phi_2)},$$

$$\tilde{A}^*\tilde{B}^* = A_0 B_0 e^{-i(2\omega t+\phi_1+\phi_2)}.$$

The time averages of the latter two terms are zero, and we are left with

$$\langle AB\rangle = \frac{1}{T}\int_0^T \frac{\tilde{A}\tilde{B}^* + \tilde{A}^*\tilde{B}}{4}\,dt.$$

Again using $\mathcal{R}e\{\tilde{z}\} = (\tilde{z}+\tilde{z}^*)/2$, we may rewrite this as

$$\langle AB\rangle = \frac{1}{2}\mathcal{R}e\left\{\tilde{A}\tilde{B}^*\right\}. \tag{2.28}$$

The reader may find this quite general relation easier to use than performing an integration such as (1.24).

2.8 Polarization

The displacement of a transverse wave is a vector quantity. We must therefore specify the frequency, phase, and direction of the wave along with the magnitude and direction of the displacement. The direction of the displacement vector is called the *direction of polarization*, and the plane containing the direction of polarization and the propagation vector is called the *plane of polarization*. This quantity has the same name as the field quantity introduced in (2.5). Because the two terms describe completely different physical phenomena, there should be no danger of confusion.

From our study of Maxwell's equations, we know that **E** and **H**, for a plane wave in free space, are mutually perpendicular and lie in a plane normal to the direction of propagation, **k**. We also know that, given one of the two vectors, we can use (2.17) to obtain the other. Convention requires that we use the electric vector to label the direction of the electromagnetic wave's polarization. The selection of the electric field is not completely arbitrary.

The electric field of the electromagnetic wave acts on a charged particle in the material with a force

$$\mathbf{F}_E = q\mathbf{E}. \tag{2.29}$$

This force accelerates the charged particle to a velocity v in a direction transverse to the direction of light propagation and parallel to the electric field. The moving charge interacts with the magnetic field of the electromagnetic wave with a force

$$\mathbf{F}_H = q(\mathbf{v}\times\mathbf{B}), \tag{2.30}$$

parallel to the propagation vector. We can write the ratio of the forces on a moving charge in an electromagnetic field due to the electric and magnetic fields as

$$\frac{F_E}{F_H} = \frac{eE}{evB}.$$

We can replace B using (2.19) to obtain

$$\frac{F_E}{F_H} = \frac{c}{nv}, \tag{2.31}$$

where v is the velocity of the moving charge. Assuming that a charged particle is traveling in air at the speed of sound, so that $v = 335$ m/s, the force due to the electric field of a light wave on that particle will therefore be 8.9×10^5 times larger than the force due to the magnetic field. The size of this numbers demonstrates that, except in relativistic situations, when $v \approx c$, the interaction of an electromagnetic wave with matter will be dominated by the electric field.

A conventional vector notation is used to describe the polarization of a light wave; however, to visualize the behavior of the electric field vector as light propagates, a geometrical construction is useful. The geometrical construction, called a Lissajous figure, describes the path followed by the tip of the electric field vector.

2.8.1 Polarization Ellipse

Assume that a plane wave is propagating in the z-direction and that the electric field, determining the direction of polarization, is oriented in the (x, y) plane. In complex notation, the plane wave is given by

$$\tilde{\mathbf{E}} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi)} = \mathbf{E}_0 e^{i(\omega t - kz + \phi)}.$$

This wave can be written in terms of the x- and y-components of $\mathbf{E}_0$:

$$\tilde{\mathbf{E}} = E_{0x} e^{i(\omega t - kz + \phi_1)}\hat{\mathbf{i}} + E_{0y} e^{i(\omega t - kz + \phi_2)}\hat{\mathbf{j}}. \tag{2.32}$$

(To prevent errors, we will use only the real part of $\mathbf{E}$ for manipulation.) We divide each component of the electric field by its maximum value so that the problem is reduced to one of the following two sinusoidally varying unit vectors:

$$\begin{aligned}
\frac{E_x}{E_{0x}} &= \cos(\omega t - kz + \phi_1) = \cos(\omega t - kz)\cos\phi_1 - \sin(\omega t - kz)\sin\phi_1, \\
\frac{E_y}{E_{0y}} &= \cos(\omega t - kz)\cos\phi_2 - \sin(\omega t - kz)\sin\phi_2.
\end{aligned}$$

When these unit vectors are added together, the result will be a set of figures called *Lissajous figures* (**Jules Antoine Lissajous: 1822–1880**). The geometrical construction shown in Figure 2.3 can be used to visualize the generation of the Lissajous figure. The harmonic motion along the x-axis is found by projecting a vector rotating around a circle of diameter E_{0x} onto the x-axis. The harmonic motion, along the y-axis is generated the same way using a circle of diameter E_{0y}. The resulting x- and y-components are added to obtain $\mathbf{E}$. In Figure 2.3, the two harmonic oscillators both have the same frequency, $\omega t - kz$, but differ in phase by

$$\delta = \phi_2 - \phi_1 = -\frac{\pi}{2}.$$

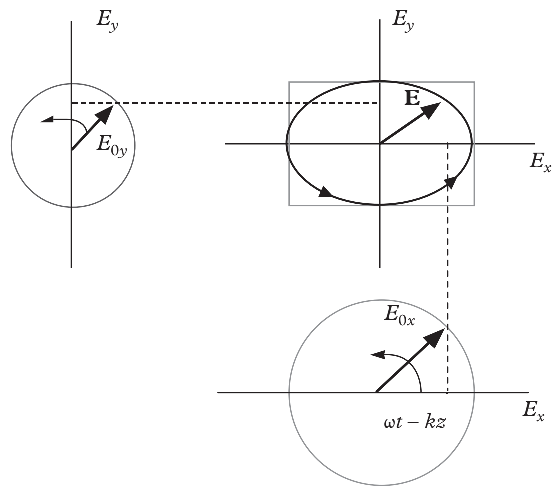

Figure 2.3 *Geometrical construction showing how the Lissajous figure is constructed from harmonic motion along the x- and y-coordinate axes. The harmonic motion along each coordinate axis is created by projecting a vector rotating around a circle onto the axis.*

The tip of the electric field **E** in Figure 2.3 traces out an ellipse, with its axes aligned with the coordinate axes. To determine the direction of the rotation of the vector, assume that $\phi_1 = 0,\ \phi_2 = -\pi/2$, and $z = 0$, so that

$$\frac{E_x}{E_{0x}} = \cos\omega t, \quad \frac{E_y}{E_{0y}} = \sin\omega t$$

$$\mathbf{E} = \left(\frac{E_x}{E_{0x}}\right)\hat{\mathbf{i}} + \left(\frac{E_y}{E_{0y}}\right)\hat{\mathbf{j}}.$$

Table 2.3 *Rotating electric field vector* **E**

ωt	**E**
0	$\hat{\mathbf{i}}$
$\frac{\pi}{4}$	$\frac{1}{\sqrt{2}}\left(\hat{\mathbf{i}}+\hat{\mathbf{j}}\right)$
$\frac{\pi}{2}$	$\hat{\mathbf{j}}$
$\frac{3\pi}{4}$	$\frac{1}{\sqrt{2}}\left(-\hat{\mathbf{i}}+\hat{\mathbf{j}}\right)$
π	$-\hat{\mathbf{i}}$

The normalized vector **E** can easily be evaluated at a number of values of ωt to discover the direction of rotation. Table 2.3 shows the value of the vector as ωt increases.

The rotation of the vector **E** in Figure 2.3 is seen to be in a counterclockwise direction, moving from the positive x-direction, to the y-direction, and finally to the negative x-direction.

To obtain the equation for the Lissajous figure, we eliminate the dependence of the unit vectors on $\omega t - kz$. First, we multiply the two equations by $\sin\phi_2$ and $\sin\phi_1$, respectively, and then subtract the resulting equations. Second, we multiply the two equations by $\cos\phi_2$ and $\cos\phi_1$, respectively, and then subtract the resulting equations. These two operations yield the following pair of equations;

$$\frac{E_x}{E_{0x}}\sin\phi_2 - \frac{E_y}{E_{0y}}\sin\phi_1 = \cos(\omega t - kz)\,[\cos\phi_1\sin\phi_2 - \sin\phi_1\cos\phi_2],$$

$$\frac{E_x}{E_{0x}}\cos\phi_2 - \frac{E_y}{E_{0y}}\cos\phi_1 = \sin(\omega t - kz)\,[\cos\phi_1\sin\phi_2 - \sin\phi_1\cos\phi_2].$$

The term in square brackets can be simplified using the trigonometric identity

$$\sin\delta = \sin(\phi_2 - \phi_1) = \cos\phi_1\sin\phi_2 - \sin\phi_1\cos\phi_2.$$

After replacing the term in square brackets by $\sin\delta$, the two equations are squared and added, yielding the equation for the Lissajous figure:

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 - \left(\frac{2E_xE_y}{E_{0x}E_{0y}}\right)\cos\delta = \sin^2\delta. \tag{2.33}$$

The trigonometric identity

$$\cos\delta = \cos(\phi_2 - \phi_1) = \cos\phi_1\cos\phi_2 + \sin\phi_1\sin\phi_2$$

was also used here to further simplify (2.33).

Equation (2.33) has the same form as the equation of a conic,

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0.$$

The conic here can be seen to be an ellipse because, from (2.33),

$$B^2 - 4AC = \frac{4}{E_{0x}^2E_{0y}^2}(\cos^2\delta - 1) < 0.$$

This ellipse is called the *polarization ellipse*. Its orientation with respect to the x-axis is given by

$$\tan 2\theta = \frac{B}{A-C} = \frac{2E_{0x}E_{0y}\cos\delta}{E_{0x}^2 - E_{0y}^2}. \tag{2.34}$$

If $A = C$ and $B \neq 0$ then $\theta = 45°$. When $\delta = \pm\pi/2$, then $\theta = 0°$, as shown in Figure 2.3.

The tip of the resultant electric field vector obtained from (2.34) traces out the polarization ellipse in the plane normal to $\mathbf{k}$, as predicted by (2.33). A generalized polarization ellipse is shown in Figure 2.4. The x- and y-coordinates of the electric field are bounded by $\pm E_{0x}$ and $\pm E_{0y}$. The rectangle in Figure 2.4 illustrates these limits. The component of the electric field along the major axis of the ellipse is

$$E_M = E_x\cos\theta + E_y\sin\theta$$

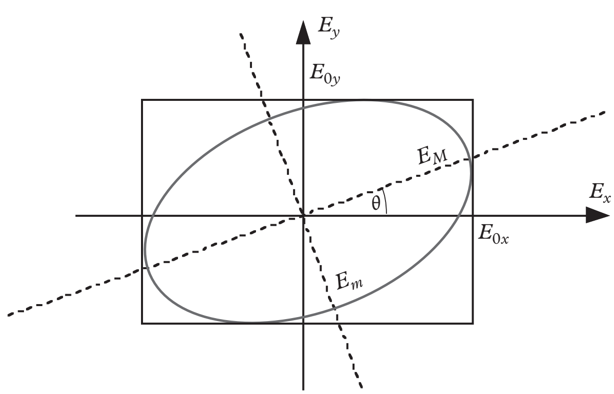

Figure 2.4 *General form of the ellipse described by (2.33).*

and that along the minor axis is

$$E_m = -E_x \sin\theta + E_y \cos\theta,$$

where θ is obtained from (2.34). The ratio of the length of the minor axis to that of the major axis is related to the ellipticity φ, which measures the amount of deviation of the ellipse from a circle:

$$\tan\varphi = \pm\left(\frac{E_m}{E_M}\right) = \frac{E_{0x}\sin\phi_1\sin\theta - E_{0y}\sin\phi_2\cos\theta}{E_{0x}\cos\phi_1\cos\theta + E_{0y}\cos\phi_2\sin\theta}. \tag{2.35}$$

To find the time dependence of the vector **E**, we rewrite (2.32) in complex form:

$$\tilde{\mathbf{E}} = e^{i(\omega t-kz)}\left(\hat{\mathbf{i}}E_{0x}e^{i\phi_1} + \hat{\mathbf{j}}E_{0y}e^{i\phi_2}\right). \tag{2.36}$$

This equation shows explicitly that the electric vector moves about the ellipse in a sinusoidal motion.

By specifying the parameters that characterize the polarization ellipse (θ and φ), we completely characterize the polarization of a wave. A review of two special cases will aid in understanding the polarization ellipse.

2.8.1.1 Linear Polarization

First consider the cases when $\delta = 0$ or π. Then (2.33) becomes

$$\left(\frac{E_x}{E_{0x}}\right)^2 + \left(\frac{E_y}{E_{0y}}\right)^2 \mp \frac{2E_xE_y}{E_{0x}E_{0y}} = 0.$$

The ellipse collapses into a straight line with slope E_{0y}/E_{0x}. The equation of the straight line is

$$\frac{E_x}{E_{0x}} = \mp\frac{E_y}{E_{0y}}.$$

Figure 2.5 displays the straight-line Lissajous figures for the two phase differences. The θ-parameter of the ellipse is the slope of the straight line,

$$\tan\theta = \frac{E_{0y}}{E_{0x}},$$

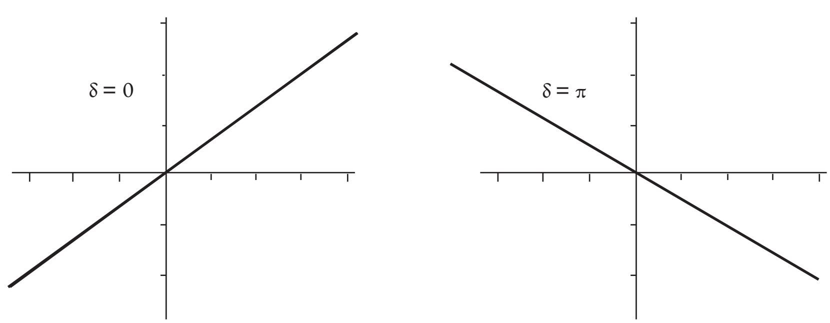

Figure 2.5 *Lissajous figures for phase differences between the y- and x-components of oscillation of 0 and π.*

resulting in the value of (2.34) being given by

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta} = \frac{2E_{0x}E_{0y}}{E_{0x}^2 - E_{0y}^2}.$$

The φ-parameter is given by (2.35) as $\tan\varphi = 0$.

The time dependence of the **E** vector shown in Figure 2.5 is given by (2.36). The real component is

$$\mathbf{E} = \left(E_{0x}\hat{\mathbf{i}} \pm E_{0y}\hat{\mathbf{j}}\right)\cos(\omega t - kz).$$

At a fixed point in space, the x- and y-components oscillate in phase (or 180° out of phase) according to the equation

$$\mathbf{E} = \left(E_{0x}\hat{\mathbf{i}} \pm E_{0y}\hat{\mathbf{j}}\right)\cos(\omega t - \phi).$$

The electric vector undergoes simple harmonic motion along the line defined by E_{0x} and E_{0y}. At a fixed time, the electric field varies sinusoidally along the propagation path (the z-axis) according to the equation

$$\mathbf{E} = \left(E_{0x}\hat{\mathbf{i}} \pm E_{0y}\hat{\mathbf{j}}\right)\cos(\phi - kz).$$

This light is said to be *linearly polarized.*

2.8.1.2 Circular Polarization

The second case occurs when $E_{0x} = E_{0y} = E_0$ and $\delta = \pm\pi/2$. From (2.33),

$$\left(\frac{E_x}{E_0}\right)^2 + \left(\frac{E_y}{E_0}\right)^2 = 1.$$

The ellipse becomes a circle as shown in Figure 2.6. For this polarization, $\tan 2\theta$ is indeterminate and $\tan\varphi = 1$.

From (2.38), the temporal behavior is given by

$$\mathbf{E} = E_0[\cos(\omega t - kz)\,\hat{\mathbf{i}} \pm \sin(\omega t - kz)\,\hat{\mathbf{j}}].$$

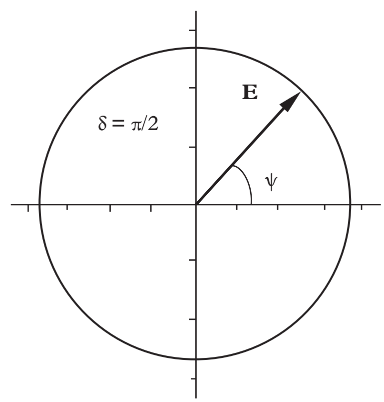

Figure 2.6 *Lissajous figure for the case when the phase differences between the y- and x-components of oscillation differ by $\pm\pi/2$ and the amplitudes of the two components are equal. The tip of the electric field vector moves along the circular path shown in the figure.*

The time dependence of the angle ψ that the **E** field makes with the x-axis in Figure 2.6 can be obtained by finding the tangent of ψ:

$$\tan\psi = \frac{E_y}{E_x} = \pm\frac{\sin(\omega t - kz)}{\cos(\omega t - kz)} = \pm\tan(\omega t - kz).$$

The interpretation of this result is that at a fixed point in space, the **E** vector rotates in a clockwise direction if $\delta = \pi/2$ and a counterclockwise direction if $\delta = -\pi/2$.

In particle physics, the light would be said to have a negative helicity if it rotated in a clockwise direction. If we look at the source, the electric vector seems to follow the threads of a left-handed screw, agreeing with the nomenclature that left-handed quantities are negative.

However, in optics, the light that rotates clockwise as we view it traveling toward us from the source is said to be *right-circularly polarized*. The counterclockwise-rotating light is *left-circularly polarized*.

The association of right-circularly polarized light with “right-handedness” in optics came about by looking at the path of the electric vector in space at a fixed time: then $\tan\psi = \tan(\phi - kz)$; see Figure 2.7. As shown in Figure 2.7, right-circularly polarized light at a fixed time seems to spiral in a counterclockwise fashion along the z-direction, following the threads of a right-handed screw.

This motion can be generalized to include elliptically polarized light when $E_{0x} \neq E_{0y}$. Figure 2.3 schematically displays the generation of the Lissajous figure for the case $\delta = \pi/2$ but with unequal values of E_{0x} and E_{0y}. Figure 2.8 shows two calculated Lissajous figures. If the electric vector moves around the ellipse in a clockwise direction, as we face the source, then the phase difference and ellipticity are

$$0 \leq \delta \leq \pi \qquad \text{and} \qquad 0 < \varphi < \frac{\pi}{4},$$

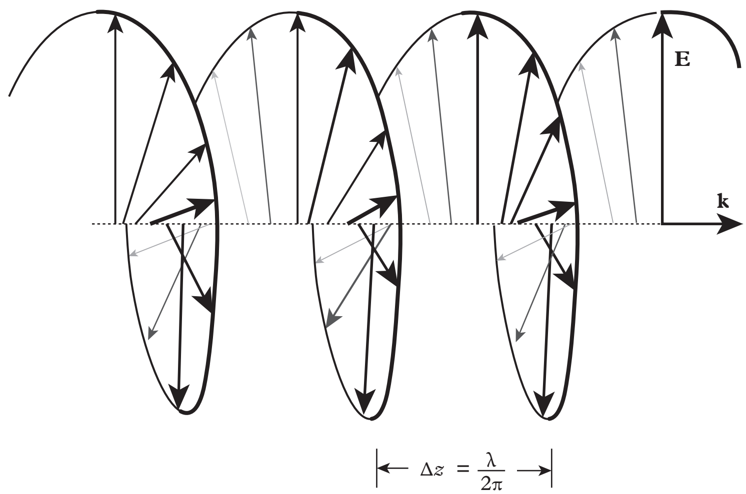

Figure 2.7 *The path of the electric vector of right-circularly polarized light at a fixed time.*

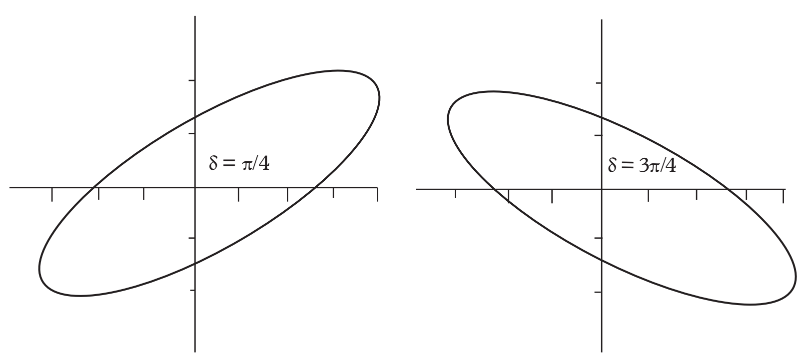

Figure 2.8 *Lissajous figures for elliptically polarized light. These were calculated with $E_{0x} = 0.75$ and $E_{0y} = 0.25$.*

and the polarization is right-handed. If the motion of the electric vector is moving in a counterclockwise direction, then the phase difference and ellipticity are

$$-\pi \leq \delta \leq 0 \qquad \text{and} \qquad -\frac{\pi}{4} < \varphi < 0.$$

The orientation of either ellipse, with respect to the x-axis, is given by (2.34) and depends upon the relative magnitudes of E_{0x} and E_{0y}.

The procedure used to decompose an arbitrary polarization into polarizations parallel to two axes of a Cartesian coordinate system is a technique used extensively in vector algebra to simplify mathematical calculations. According to the mathematical formalism associated with this technique, the polarization is described in terms of a set of basis vectors $\mathbf{e}_i$. An arbitrary polarization would be expressed as

$$\mathbf{E} = \sum_{i=1}^{2} a_i \mathbf{e}_i. \tag{2.37}$$

The set of basis vectors $\mathbf{e}_i$ are orthonormal, i.e.,

$$\mathbf{e}_i \mathbf{e}_j^* = \delta_{ij} = \begin{cases} 1 & (i = j), \\ 0 & (i \neq j), \end{cases}$$

where we have assumed that the basis vectors could be complex. We mention this mathematical formalism because an identical formalism is encountered in elementary particle physics, where it is used to describe spin [4].

In a Cartesian coordinate system, the $\mathbf{e}_i$'s are the unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$. The summation in (2.36) extends over only two terms because the electromagnetic wave is transverse, confining **E** to a plane normal to the direction of propagation [according to the coordinate convention we have selected, the **E** field is in the (x, y) plane].

The polarization can also be described in terms of basis vectors consisting of a right-circularly polarized component

$$\mathbf{E}_{\mathcal{R}} = E_{0\mathcal{R}}[\cos(\omega t - kz)\,\hat{\mathbf{i}} - \sin(\omega t - kz)\,\hat{\mathbf{j}}]$$

and a left-circularly polarized component

$$\mathbf{E}_{\mathcal{L}} = E_{0\mathcal{L}}[\cos(\omega t - kz)\,\hat{\mathbf{i}} + \sin(\omega t - kz)\,\hat{\mathbf{j}}].$$

An arbitrary elliptical polarization can then be written as

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_{\mathcal{R}} + \mathbf{E}_{\mathcal{L}} \\ &= (E_{0\mathcal{R}} + E_{0\mathcal{L}}) \cos(\omega t - kz)\,\hat{\mathbf{i}} - (E_{0\mathcal{R}} - E_{0\mathcal{L}}) \sin(\omega t - kz)\,\hat{\mathbf{j}}. \end{aligned} \tag{2.38}$$

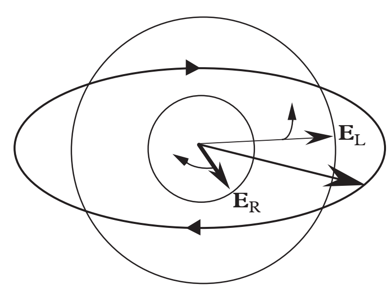

Figure 2.9 *Construction of elliptically polarized light from two circularly polarized waves.*

The geometrical construction that demonstrates the expression of an arbitrary elliptically polarized light wave in terms of right- and left-circularly polarized waves is shown in Figure 2.9. The use of circularly polarized waves as the basis set for describing polarization is discussed by Klein and Furtak [5].

In the formalism associated with (2.37), the expansion coefficients a_i can be used to form a 2×2 matrix that in statistical mechanics is called the density matrix and in optics the coherency matrix [6]. The elements of this matrix are formed by the rule

$$\rho_{ij} = \mathbf{a}_i \mathbf{a}_j^*.$$

We will not develop the theory of polarization using the coherency matrix, but will simply use this matrix to justify the need for four independent measurements to characterize polarization. There is no unique set of measurements required by theory, but normally measurements made are of the *Stokes parameters*, which are directly related to the polarization ellipse of Figure 2.4. (We will see shortly that only three of the four measurements are independent. This will be in agreement with the definition of the coherency matrix, where $\rho_{ij} = \rho_{ij}^*$, i.e., the matrix is Hermitian.)

2.8.2 Stokes Parameters

The Stokes parameters (**Sir George Gabriel Stokes: 1819–1903**) of a light wave are measurable quantities, defined as follows:

s_0 total flux density;

s_1 difference between flux densities transmitted by a linear polarizer oriented parallel to the x-axis and one oriented parallel to the y-axis (the x- and y-axes are usually selected to be parallel to the horizontal and vertical directions in the laboratory);

s_2 difference between flux densities transmitted by a linear polarizer oriented at 45° to the x-axis and one oriented at 135°;

s_3 difference between flux densities transmitted by a right-circular polarizer and a left-circular polarizer.

The physical instruments that can be used to measure the parameters will be discussed in Chapter 13.

If the Stokes parameters are to characterize the polarization of a wave, they must be related to the parameters of the polarization ellipse. It is therefore important to establish that the Stokes parameters are variables of the polarization ellipse (2.33). In its current form, (2.33) contains no measurable quantities and thus must be modified if it is to be associated with the Stokes parameters. In the discussion of the Poynting vector, it was pointed out that the time average of the Poynting vector is the quantity observed when measurements are made of light waves. We must, therefore, find the time average of (2.33) if we wish to relate its parameters to observable quantities. To simplify the discussion, let us assume that the amplitudes of the orthogonally polarized waves E_{0x} and E_{0y} and their relative phase δ are constants. We will also use the shorthand notation for a time average introduced in (2.24):

$$\langle E_x^2 \rangle = \frac{1}{T} \int_{t_0}^{t_0+T} E_{0x}^2 [\cos(\omega t - kz) \cos\phi_1 - \sin(\omega t - kz) \sin\phi_1]^2 \, dt.$$

The time average of (2.33) can now be written as

$$\frac{\langle E_x^2 \rangle}{E_{0x}^2} + \frac{\langle E_y^2 \rangle}{E_{0y}^2} - 2\frac{\langle E_x E_y \rangle}{E_{0x} E_{0y}} \cos\delta = \sin^2\delta. \tag{2.39}$$

Multiplying both sides of (2.39) by $(2E_{0x}E_{0y})^2$ removes the terms in the denominators to give

$$4E_{0y}^2\left\langle E_x^2\right\rangle + 4E_{0x}^2\left\langle E_y^2\right\rangle - 8E_{0x}E_{0y}\left\langle E_xE_y\right\rangle\cos\delta = (2E_{0x}E_{0y}\sin\delta)^2.$$

The same argument that was used to simplify (2.25) can be used to obtain the time averages for the first two terms:

$$\left\langle E_x^2\right\rangle = \frac{E_{0x}^2}{2}, \qquad \left\langle E_y^2\right\rangle = \frac{E_{0y}^2}{2}.$$

The calculation of the time average in the third term,

$$\left\langle E_xE_y\right\rangle = \frac{1}{2}E_{0x}E_{0y}\cos\delta, \tag{2.40}$$

is left as Problem 2.12. With these time averages, (2.39) can be written as

$$4E_{0x}^2E_{0y}^2 - \left(2E_{0x}E_{0y}\cos\delta\right)^2 = \left(2E_{0x}E_{0y}\sin\delta\right)^2.$$

If $E_{0x}^2 + E_{0y}^2$ is added to both sides of this equation, it can be rewritten as

$$\left(E_{0x}^2 + E_{0y}^2\right)^2 - \left(E_{0x}^2 - E_{0y}^2\right)^2 - \left(2E_{0x}E_{0y}\cos\delta\right)^2 = \left(2E_{0x}E_{0y}\sin\delta\right)^2. \tag{2.41}$$

Each term in this equation can be identified with a Stokes parameter.

In our derivation we required that the amplitudes and relative phase of the two orthogonally polarized waves be constant, but we can relax this requirement and instead define the Stokes parameters as temporal averages. With this modification, the terms of (2.41) become

$$\begin{aligned} s_0 &= \left\langle E_{0x}^2\right\rangle + \left\langle E_{0y}^2\right\rangle, \quad & s_1 &= \left\langle E_{0x}^2\right\rangle - \left\langle E_{0y}^2\right\rangle, \\ s_2 &= \left\langle 2E_{0x}E_{0y}\cos\delta\right\rangle, \quad & s_3 &= \left\langle 2E_{0x}E_{0y}\sin\delta\right\rangle. \end{aligned} \tag{2.42}$$

Equation (2.41) can now be written as

$$s_0^2 - s_1^2 - s_2^2 = s_3^2. \tag{2.43}$$

For a polarized wave, only three of the Stokes parameters are independent. This agrees with the requirement placed upon elements of the Hermitian coherency matrix, introduced above.

With this demonstration of the connection between the Stokes parameters and the polarization ellipse, the Stokes parameters can be written in terms of the parameters of the polarization ellipse in Figure 2.4:

$$\begin{aligned} s_1 &= s_0\cos 2\varphi\cos 2\theta, \\ s_2 &= s_0\cos 2\varphi\sin 2\theta, \\ s_3 &= s_0\sin 2\varphi. \end{aligned} \tag{2.44}$$

It is this close relationship between the Stokes parameters and the polarization ellipse that makes the Stokes parameters a useful characterization of polarization.

The Stokes parameters can be used to define the *degree of polarization*

$$V = \frac{1}{s_0}\sqrt{s_1^2 + s_2^2 + s_3^2}. \tag{2.45}$$

[The equality (2.43) applies to completely polarized light, when $V = 1$.] The degree of polarization can be used to characterize any light source that is physically realizable. If the time averages in the definition of the Stokes parameters s_2 and s_3 in (2.42) are zero and if $\langle E_{0x}^2 \rangle = \left\langle E_{0y}^2 \right\rangle$, so that

$$s_0 = 2\langle E_{0x}^2 \rangle, \quad s_1 = s_2 = s_3 = 0,$$

then the light wave is said to be unpolarized and $V = 0$.

Hans Mueller, a physics professor at MIT [6], pointed out that the Stokes parameters can be thought of as elements of a column matrix or 4-vector

$$\begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{bmatrix}.$$

This view will allow us to follow a polarized wave through a series of optical devices by using matrix algebra, as we will see later.

Poincaré's sphere (Henri Poincaré: 1854–1912) Before it was discovered that the Stokes parameters could be treated as elements of a column matrix, a geometrical construction was used to determine the effect of an anisotropic medium on polarized light. The parameters s_1, s_2, s_3 are viewed as the Cartesian coordinates of a point on a sphere of radius s_0. This sphere is called the *Poincaré sphere* [7] and is shown in Figure 2.10. On the sphere, right-hand polarized light is represented by points on the upper hemisphere. Linear polarization is represented by points on the equator. Circular polarization is represented by the poles.

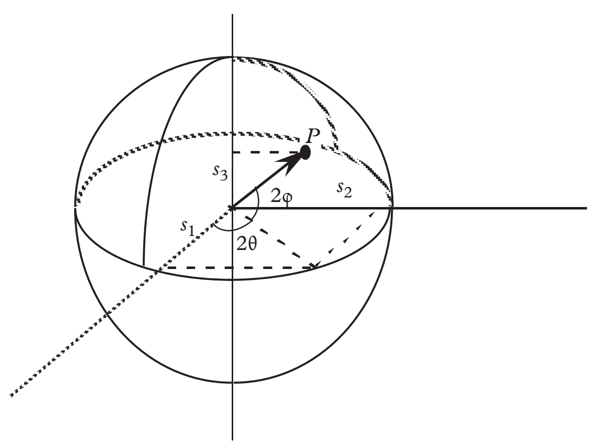

Figure 2.10 *Poincaré sphere.*

2.8.3 Jones Vector

There is one other representation of polarized light, complementary to the Stokes parameters, developed by **R. Clark Jones (1916–2004)** in 1941 and called the *Jones vector*. It is superior to the Stokes vector in that it handles light of a known phase and amplitude with a reduced number of parameters. It is inferior to the Stokes vector in that, unlike the Stokes representation, which is experimentally determined, the Jones representation cannot handle unpolarized or partially polarized light. The Jones vector is a theoretical construct that can only describe light with a well-defined phase and frequency. The density matrix formalism can be used to correct the shortcomings of the Jones vector, but then the simplicity of the Jones representation is lost.

It was shown earlier that if it is assumed that the coordinate system is such that the electromagnetic wave is propagating along the z-axis, then any polarization can be decomposed

Table 2.4 *Jones and Stokes vectors*

Jones vector	**Stokes vector**
Horizontal polarization	
$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\1\\0\\0\end{bmatrix}$
Vertical polarization	
$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\-1\\0\\0\end{bmatrix}$
+45° polarization	
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$
–45° polarization	
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$	$\begin{bmatrix}1\\0\\-1\\0\end{bmatrix}$
Right-circular polarization	
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$	$\begin{bmatrix}1\\0\\0\\1\end{bmatrix}$
Left-circular polarization	
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$	$\begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$

into two orthogonal **E** vectors, say, for the purposes of this discussion, parallel to the x- and y-directions. The Jones vector is defined as a two-row column matrix consisting of the complex components in the x- and y-directions:

$$\mathbf{E} = \begin{bmatrix} E_{0x} e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_1)} \\ E_{0y} e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_2)} \end{bmatrix}. \tag{2.46}$$

If absolute phase is not an issue, then we may normalize this vector by dividing it by that number that simplifies the components but keeps the sum of the squared magnitudes of the components equal to one. For example, assuming that $E_{0x} = E_{0y}$, then

$$\mathbf{E} = E_{0x} e^{i(\omega t - \mathbf{k}\cdot\mathbf{r} + \phi_1)} \begin{bmatrix} 1 \\ e^{i\delta} \end{bmatrix}.$$

The normalized vector would comprise the terms contained within the square brackets, each divided by $\sqrt{2}$. The general form of the Jones vector is

$$\mathbf{E} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}, \qquad \mathbf{E}^* = \begin{bmatrix} \mathbf{A}^* & \mathbf{B}^* \end{bmatrix}.$$

Some examples of Jones vectors and Stokes vectors are shown in Table 2.4.

2.9 Propagation in a Conducting Medium

In Chapter 1, we discussed the propagation of a wave with attenuation. In our discussion of the propagation of light, however, we have ensured that we would experience no loss by assuming $\sigma = 0$. We now relax that assumption and allow $\sigma \neq 0$. Maxwell's equations become

$$\nabla \cdot \mathbf{D} = 0, \qquad \nabla \cdot \mathbf{B} = 0,$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}.$$

We continue to neglect dynamic or resonant effects so that we may use the simple constitutive relations

$$\mathbf{J} = \sigma \mathbf{E}, \quad \mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}.$$

where ε, μ, and σ are scalars, independent of time. Maxwell's equations in a medium with dissipation can be rewritten using these constitutive relations as

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0, & \nabla \cdot \mathbf{H} &= 0, \\ \nabla \times \mathbf{H} &= \sigma \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t}, & \nabla \times \mathbf{E} &= -\mu \frac{\partial \mathbf{H}}{\partial t}. \end{aligned} \tag{2.47}$$

We now apply the same procedure used to derive the wave equation for free space,

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\mu \frac{\partial \mathbf{H}}{\partial t}\right) = -\mu \frac{\partial}{\partial t}(\nabla \times \mathbf{H}),$$

$$-\mu \frac{\partial}{\partial t}(\nabla \times \mathbf{H}) = -\mu \frac{\partial}{\partial t}\left(\sigma \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t}\right),$$

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}.$$

yielding the wave equation in a conducting medium:

$$\nabla^2 \mathbf{E} = \mu\sigma \frac{\partial \mathbf{E}}{\partial t} + \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{2.48}$$

This wave equation is of the same form as (1.20). We can derive a similar equation for the magnetic field:

$$\nabla^2 \mathbf{B} = \mu\sigma \frac{\partial \mathbf{B}}{\partial t} + \mu\varepsilon \frac{\partial^2 \mathbf{B}}{\partial t^2}. \tag{2.49}$$

Equations (2.48) and (2.49) are called the *telegraph equations* and were developed by Oliver Heaviside (1850–1925). They are wave equations derived to explain the propagation of pulses on telegraph lines.

We see that the wave equation (2.48) contains a damping term $\partial \mathbf{E}/\partial t$ when we allow $\sigma \neq 0$. By comparing (2.48) and (1.20), we can see that the solution of (2.48) will be an electromagnetic wave that will experience attenuation proportional to $\mu\sigma$ as it propagates. Using (1.22) and (1.23), we can rewrite (2.47), for plane wave solutions, as

$$\begin{aligned} \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad &\Rightarrow \quad \nabla \times \mathbf{E} = -i\omega\mu \mathbf{H}, \\ \nabla \times \mathbf{H} = \sigma \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \quad &\Rightarrow \quad \nabla \times \mathbf{H} = i\omega\left(\varepsilon - \frac{i\sigma}{\omega}\right)\mathbf{E}. \end{aligned} \tag{2.50}$$

We can rewrite (2.48) in terms of these expressions for the curls of **E** and **H**:

$$\nabla^2 \mathbf{E} + \omega^2 \mu \left(\varepsilon - \frac{i\sigma}{\omega}\right)\mathbf{E} = 0. \tag{2.51}$$

This has the form of the Helmholtz equation (1.13) if we replace k^2 in the latter by the complex function

$$\tilde{k}^2 = \omega^2 \mu \left(\varepsilon - i\frac{\sigma}{\omega}\right). \tag{2.52}$$

We use the identity

$$k = \frac{n\omega}{c} = \omega\sqrt{\mu\varepsilon},$$

to demonstrate that the equations for conducting media are identical to those derived for nonconducting media if the dielectric constant ε is replaced by a complex dielectric constant

$$\tilde{\varepsilon} = \varepsilon - i\left(\frac{\sigma}{\omega}\right). \tag{2.53}$$

This equation suggests that σ may contain a frequency dependence (in fact, in the c.g.s. system, the units of σ are s^{-1}; for copper, in c.g.s. units, $\sigma = 5.14 \times 10^{17}\ s^{-1}$). In condensed matter physics, one finds that the mobility of the electrons creates a frequency dependence that shows up in σ.

Since we have replaced k by the complex quantity

$$\tilde{k} = \omega\sqrt{\mu\left(\varepsilon - i\frac{\sigma}{\omega}\right)},$$

we must replace the index of refraction by a complex index. In the literature, this is accomplished in two ways:

$$\tilde{n} = n(1 - i\kappa), \tag{2.54}$$

$$\tilde{n} = n_1 - in_2.$$

We will use the notation in (2.54).

To find out how a plane wave propagates in this conductive medium, we simply replace the propagation constant k by

$$\tilde{k} = \tilde{n}\frac{\omega}{c} = \left(\frac{\omega n}{c}\right)(1 - i\kappa),$$

as we did with (1.24). κ is called the *extinction coefficient* and $n\kappa$ is called the *absorption coefficient*.

If we assume that $\mathbf{k}$ is parallel to the z-axis, then the plane wave is

$$\mathbf{E} = \mathbf{E}_0 e^{i\omega t} e^{-i\omega(n/c)(1-i\kappa)z},$$

$$\mathbf{E} = \mathbf{E}_0 e^{-(\omega n\kappa/c)z} e^{-i\omega(t-nz/c)}, \tag{2.55}$$

$$\mathcal{Re}\{\mathbf{E}\} = \mathbf{E}_0 e^{-(\omega n\kappa/c)z} \cos(\omega t - kz). \tag{2.56}$$

The wave described by (2.56) is a plane wave, attenuated by the exponential factor

$$e^{-(\omega/c)n\kappa z}. \tag{2.57}$$

Figure 2.11 *Blue laser light is shown propagating in xylene (above) and water (below). The water contains the dye rhodamine 6G in solution. The red rhodamine dye absorbs the blue light, and the beam rapidly decays to zero. Some of the energy absorbed by the dye is reemitted in the yellow to red region of the spectrum. This reemitted light caused the diffuse appearance of the light as it propagates in the water.*

Figure 2.11 displays the exponential decay of a light wave propagating in an absorbing medium. A layer of xylene floats upon water containing the dye rhodamine 6G in solution. The dye strongly absorbes a beam of blue light from a HeCd laser (440 nm). As can be seen in Figure 2.11, the blue light is rapidly attenuated once it enters the water. Some of the energy absorbed by the rhodamine is reemitted at longer wavelengths. The reemitted light travels in all directions, since it has no memory of the direction traveled by the blue light. For this reason, the beam of light in the water appears diffuse, and as can be seen, is orange in color.

To evaluate the absorption coefficient $n\kappa$ in terms of electromagnetic properties of the medium, we will derive a relationship between $n\kappa$ and σ. We rewrite (2.54) as

$$\tilde{n}^2 = n^2(1-\kappa^2-2i\kappa) = \frac{c^2}{\omega^2}\tilde{k}^2.$$

This can be used to express n^2 in terms of the constants of the material:

$$\tilde{n}^2 = c^2\mu\left(\varepsilon - i\frac{\sigma}{\omega}\right). \tag{2.58}$$

Equating real and imaginary terms, we obtain

$$n^2(1-\kappa^2) = c^2\mu\varepsilon, \quad 2n^2\kappa = c^2\frac{\mu\sigma}{\omega}.$$

We can use these two relationships to find

$$n^2 = \frac{c^2}{2}\left[\sqrt{\mu^2\varepsilon^2+\left(\frac{\mu\sigma}{\omega}\right)^2}+\mu\varepsilon\right], \tag{2.59}$$

$$n^2\kappa^2 = \frac{c^2}{2}\left[\sqrt{\mu^2\varepsilon^2+\left(\frac{\mu\sigma}{\omega}\right)^2}-\mu\varepsilon\right]. \tag{2.60}$$

Note that when $\sigma = 0$, $\kappa = 0$, and we obtain the free-space result, (2.11):

$$n^2 = \frac{\mu\varepsilon}{\mu_0\varepsilon_0}.$$

An estimate of the magnitude of the quantities under the radicals in (2.59) and (2.60) can be obtained by using values for copper, where, in MKS units, $\sigma = 5.8 \times 10^7$ mho/m and $n = 0.62$ at $\lambda = 589.3$ nm. (The index of refraction is less than one, which implies that the phase velocity is greater than the speed of light. This apparent contradiction of a fundamental postulate of the theory of relativity will be discussed during the study of dispersion in Chapter 7.) The two terms under the radical are

$$\frac{\mu\sigma}{\omega} = \frac{\left(4\pi\times10^{-7}\right)\left(5.8\times10^{7}\right)\left(5.893\times10^{-7}\right)}{(2\pi)\left(3\times10^{8}\right)} = 2.3\times10^{-14}\ \mathrm{s^2/m^2},$$

$$\mu\varepsilon = \mu_0\varepsilon_0 n^2 = \left(4\pi\times10^{-7}\right)\left(8.8542\times10^{-12}\right)(0.62)^2 = 4.3\times10^{-18}\ \mathrm{s^2/m^2}.$$

By comparing the relative magnitude of these two terms, we are justified in assuming that $\sigma/\omega \gg \varepsilon$ and can make the approximation

$$n^2\kappa^2 \approx \frac{c^2\mu\sigma}{2\omega},$$

$$n\kappa = c\sqrt{\frac{\mu\sigma}{2\omega}}. \tag{2.61}$$

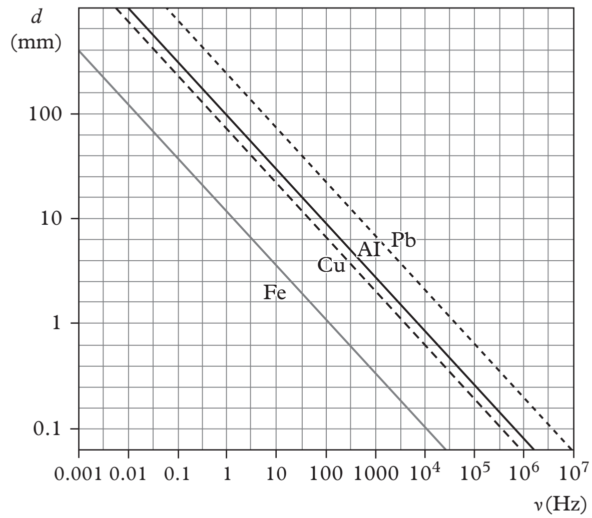

Figure 2.12 *Skin depth for a few metals.*

We use (2.61) to find the depth at which an electromagnetic wave is attenuated, to $1/e$ of its original energy, when propagating into a conductor. At that depth, denoted by d, the exponent in (2.57) will equal 1, and thus

$$\frac{\omega}{c} n\kappa d = \frac{2\pi}{\lambda_0} n\kappa d = 1,$$

$$d = \frac{\lambda_0}{2\pi n\kappa} \approx \frac{\lambda_0}{2\pi c}\sqrt{\frac{2\omega}{\mu\sigma}},$$

$$d = \sqrt{\frac{2}{\mu\sigma\omega}}. \tag{2.62}$$

The depth d is called the *skin depth*. Figure 2.12 shows the dependence of skin depth on frequency for some metals.

2.10 Summary

In this chapter, Maxwell's equations were used to obtain the wave equation for free space,

$$\nabla^2\mathbf{E} = \mu\varepsilon\frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

and for a conductive medium,

$$\nabla^2\mathbf{E} = \mu\sigma\frac{\partial \mathbf{E}}{\partial t} + \mu\varepsilon\frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

While the forms of these two equations appear quite different, we demonstrated that plane wave solutions existed for both equations when the dielectric constant of the conductive medium was replaced by a complex constant

$$\tilde{\varepsilon} = \varepsilon - i\left(\frac{\sigma}{\omega}\right).$$

This replacement means that the optical properties of a conductive material are described by a complex index of refraction

$$\tilde{n} = n(1 - i\kappa).$$

A plane wave propagating in conductive material has an amplitude attenuated by the exponential factor

$$e^{-(\omega/c)n\kappa z}.$$

By manipulation of Maxwell's equations, we were able to show that the propagation velocity of a light wave is governed by the electrical properties of the medium. The index of refraction was used to indicate the propagation velocity in the medium, relative to the propagation velocity in a vacuum:

$$n = \frac{c}{v} = \sqrt{\frac{\varepsilon\mu}{\varepsilon_0\mu_0}}.$$

Also, we were able to demonstrate that light waves must be transverse waves and that the magnitudes of the electric and magnetic fields are related by

$$n|\mathbf{E}| = c|\mathbf{B}|.$$

By comparing the forces experienced by a charged particle in an electromagnetic field, we found that we could describe the polarization of an electromagnetic wave by the electric field vector. We developed the formalism necessary for discussion of polarization, but delayed a discussion of the manipulation of a light wave's polarization until Chapter 13.

2.11 Problems

2.1. Light is traveling in glass ($n = 1.5$). If the amplitude of the electric field of the light is 100 V/m, what is the amplitude of the magnetic field? What is the magnitude of the Poynting vector?

2.2. A 60 W monochromatic point source is radiating equally in all directions in a vacuum. What is the electric field amplitude 2 m from the source?

2.3. The flux density at the Earth's surface due to sunlight is $I = 1.34 \times 10^3$ J/(m^2·s). Calculate the electric and magnetic fields at the Earth's surface by assuming that the average Poynting vector is equal to this flux density.

2.4. What is the flux density of light needed to keep a glass sphere of mass 10^{-8} g and diameter 2×10^{-5} m floating in midair?

2.5. An 85 kg astronaut has only a flashlight to propel him in space. If the flashlight emits 1 W of light in a parallel beam for one hour, how fast will the astronaut be going at the end of the hour, assuming he started at rest?

2.6. What are the polarizations of the following waves:

$$\mathbf{E} = E_0\left[\cos(\omega t - kz)\,\hat{\mathbf{i}} + \cos\left(\omega t - kz + \frac{5}{4}\pi\right)\hat{\mathbf{j}}\right],$$

$$\mathbf{E} = E_0\left[\cos(\omega t + kz)\,\hat{\mathbf{i}} + \cos\left(\omega t + kz - \frac{1}{4}\pi\right)\hat{\mathbf{j}}\right],$$

$$\mathbf{E} = E_0\left[\cos(\omega t - kz)\,\hat{\mathbf{i}} - \cos\left(\omega t - kz + \frac{1}{6}\pi\right)\hat{\mathbf{j}}\right]?$$

2.7. Show that the addition of two elliptically polarized waves propagating along the z-axis results in another elliptically polarized wave.

2.8. Write an expression, in MKS units, for a plane electromagnetic wave, with a wavelength of 500 nm and an intensity of 53.2 W/m^2, propagating in the z-direction. Assume that the wave is linearly polarized at an angle of 45° to the x-axis.

2.9. Using conventional vector notation, prove that a right- and a left-circularly polarized wave can be combined to yield a linearly polarized wave. Carry out the same demonstration using the Jones vector notation. What requirement must be placed on the two circularly polarized waves? Sketch the geometrical construction that demonstrates the combination of circularly polarized waves to generate a linearly polarized wave.

2.10. Write the equation for a plane wave propagating in the positive z-direction that has right elliptical polarization with the major axis of the ellipse parallel to the x-axis. Use both the conventional vector and the Jones vector notation.

2.11. Describe the polarization of a wave with the Jones vector

$$\begin{bmatrix} -i \\ 2 \end{bmatrix}.$$

Write the Jones vector that is orthogonal to this vector and describe its polarization.

2.12. Prove that (2.42) is correct using the expressions

$$E_x = E_{0x}[\cos(\omega t - kz)\cos\phi_1 - \sin(\omega t - kz)\sin\phi_1],$$

$$E_y = E_{0y}[\cos(\omega t - kz)\cos\phi_2 - \sin(\omega t - kz)\sin\phi_2]$$

for the two orthogonally polarized electric fields.

2.13. Demonstrate, using the Jones vector notation, that right- and left-circularly polarized light waves are orthogonal.

2.14. Find the skin depth for seawater with a resistivity $\rho = 0.20\ \Omega$/m for $\nu = 30$ kHz and 30 MHz. What frequency should we use to communicate with a submarine that will not be deeper than 100 m?

2.15. At what frequency would the approximation used to obtain (2.26) produce a 10% error?

2.16. If a 1 kW laser beam is focused to a spot with an area of 10^{-9} m^2, what is the amplitude of the electric field at the focus?

2.17. The human eye is sensitive to light of wavelengths from approximately 600 nm (red) to 400 nm (blue). (a) Calculate the frequency of both wavelengths. (b) Find the energy of the photons associated with the red and blue wavelength limits.

2.18. If green light, 500 nm, could be frequency modulated to 0.1% of the light wave's frequency, calculate the number of 6 MHz bandwidth TV channels that could be carried by the modulation.

2.19. Given the Stokes vector

$$\begin{bmatrix} 1 \\ 0 \\ -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix},$$

(a) calculate the degree of polarization, (b) determine the orthogonal vector, and (c) draw the polarization ellipse.

2.20. How thin must a sheet of iron be if it is one skin depth thick? How many atoms thick is such a sheet?

REFERENCES

1. Maxwell, J.C., *A dynamical theory of the electromagnetic field*. Phil. Trans. R. Soc. Lond., 1865. **155**: p. 459–512.
2. Feynman, R.P., *QED, The Strange Theory of Light and Matter*. 2006: Princeton University Press: p. 192.
3. Wangsness, R.K., *Electromagnetic Fields*. 2nd ed. 1986: Wiley.
4. McMaster, W.H., *Polarization and the Stokes parameters*. Am. J. Phys., 1954. **22**: p. 351–362.
5. Klein, M.V. and T.E. Furtak, *Optics*. 2nd ed. 1986: Wiley.
6. Mueller, H., *The foundation of optics*. J. Opt. Soc. Am, 1948. **38**: p. 661.
7. Poincaré, H., *Théorie mathématique de la lumiére*. Vol. 2. 1892: Georges Carré.

Appendix 2A Vectors

We will review a few properties of vectors that will be of use in our discussion of light. A vector is a quantity with both magnitude and direction; it can be defined in terms of unit vectors along the three orthogonal axes of a Cartesian coordinate system:

$$\mathbf{E} = E_x\hat{\mathbf{i}} + E_y\hat{\mathbf{j}} + E_z\hat{\mathbf{k}}.$$

To add two vectors, we add like components:

$$\mathbf{E}_1 + \mathbf{E}_2 = (E_{1x} + E_{2x})\hat{\mathbf{i}} + (E_{1y} + E_{2y})\hat{\mathbf{j}} + (E_{1z} + E_{2z})\hat{\mathbf{k}}.$$

2A.1 Products

There are two ways to multiply vectors:

- The *scalar (dot) product* is given by

$$\mathbf{E} \cdot \mathbf{H} = EH\cos\theta = E_x H_x + E_y H_y + E_z H_z, \tag{2A.1}$$

where θ is the angle between $\mathbf{E}$ and $\mathbf{H}$. This product is a scalar quantity that gives the projection of one vector onto the second vector. If $\mathbf{E}$ and $\mathbf{H}$ are perpendicular, then their dot product is zero.

- The *vector (cross) product* is given by

$$\mathbf{E} \times \mathbf{H} = EH\sin\theta\, \hat{\mathbf{n}},$$

where $\hat{\mathbf{n}}$ is a unit vector normal to the plane formed by $\mathbf{E}$ and $\mathbf{H}$. The cross product is a vector with a magnitude equal to the area of the parallelogram formed by $\mathbf{E}$ and $\mathbf{H}$; it is zero if the two vectors are parallel. The components of this vector are calculated as follows:

$$\begin{aligned}\mathbf{E} \times \mathbf{H} &= \begin{bmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ E_x & E_y & E_z \\ H_x & H_y & H_z \end{bmatrix} \\ &= (E_y H_z - E_z H_y)\hat{\mathbf{i}} - (E_x H_z - E_z H_x)\hat{\mathbf{j}} + (E_x H_y - E_y H_x)\hat{\mathbf{k}}.\end{aligned} \tag{2A.2}$$

We will find use for the vector triple product relationship

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}. \tag{2A.3}$$

2A.2 Derivatives

A vector operator called the *del operator* is defined in a Cartesian coordinate system as

$$\nabla = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}. \tag{2A.4}$$

We can treat this operator as a vector and calculate three products of use in optics:

- The *gradient* of a scalar:

$$\nabla V = \frac{\partial V}{\partial x}\hat{\mathbf{i}} + \frac{\partial V}{\partial y}\hat{\mathbf{j}} + \frac{\partial V}{\partial z}\hat{\mathbf{k}}. \tag{2A.5}$$

The gradient is a vector giving the magnitude and direction of the fastest rate of change of the scalar quantity V.

- The *divergence* of a vector:

$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}. \tag{2A.6}$$

The divergence gives the amount of flux flowing toward (negative) or away from (positive) a point. If the divergence is zero, then there are no sources or sinks in the volume.

- The *curl* of a vector:

$$\nabla \times \mathbf{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\hat{\mathbf{i}} + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\hat{\mathbf{j}} + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\hat{\mathbf{k}}. \tag{2A.7}$$

A physical interpretation of this operation can be made easily if the vector is a velocity, then when the curl of the velocity is nonzero, rotation is also occurring.

If we calculate the divergence of a gradient of a scalar, we obtain a scalar function, the *Laplacian*:

$$\nabla \cdot \nabla V = \nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \tag{2A.8}$$

For a vector quantity, we have

$$\nabla^2 \mathbf{E} = \nabla^2 E_x \,\hat{\mathbf{i}} + \nabla^2 E_y \,\hat{\mathbf{j}} + \nabla^2 E_z \,\hat{\mathbf{k}}. \tag{2A.9}$$

Table 2A.1 *Operations with the del operator in spherical coordinates (r, θ, ϕ), where subscripts r, θ, and ϕ denote the respective components*

$$(\nabla V)_r = \frac{\partial V}{\partial r} \qquad (\nabla V)_\theta = \frac{1}{r}\left(\frac{\partial V}{\partial \theta}\right) \qquad (\nabla V)_\phi = \frac{1}{r\sin\theta}\left(\frac{\partial V}{\partial \phi}\right)$$

$$\nabla \cdot \mathbf{E} = \frac{1}{r}\left[\frac{\partial (rE_r)}{\partial r}\right] + \frac{1}{r\sin\theta}\left[\frac{\partial (E_\theta \sin\theta)}{\partial \theta}\right] + \frac{1}{r\sin\theta}\left(\frac{\partial E_\phi}{\partial \phi}\right)$$

$$\nabla^2 V = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial V}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 V}{\partial \phi^2}$$

$$(\nabla \times \mathbf{E})_r = \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial \theta}(E_\phi \sin\theta) - \frac{\partial E_\theta}{\partial \phi}\right]$$

$$(\nabla \times \mathbf{E})_\theta = \frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial E_r}{\partial \phi} - \frac{\partial}{\partial r}(rE_\phi)\right]$$

$$(\nabla \times \mathbf{E})_\phi = \frac{1}{r}\left[\frac{\partial}{\partial r}(rE_\theta) - \frac{\partial E_r}{\partial \theta}\right]$$

$$(\nabla^2 \mathbf{E})_r = \nabla^2 E_r - \frac{2E_r}{r^2} - \frac{2}{r^2 \sin\theta}\frac{\partial (E_\theta \sin\theta)}{\partial \theta} - \frac{2}{r^2 \sin\theta}\frac{\partial E_\phi}{\partial \phi}$$

$$(\nabla^2 \mathbf{E})_\theta = \nabla^2 E_\theta + \frac{2}{r^2}\frac{\partial E_r}{\partial \theta} - \frac{E_\theta}{r^2 \sin^2\theta} - \frac{2\cos\theta}{r^2 \sin^2\theta}\frac{\partial E_\phi}{\partial \phi}$$

$$(\nabla^2 \mathbf{E})_\phi = \nabla^2 E_\phi + \frac{2}{r^2 \sin\theta}\frac{\partial E_r}{\partial \phi} + \frac{2\cos\theta}{r^2 \sin^2\theta}\frac{\partial E_\theta}{\partial \phi} - \frac{E_\phi}{r^2 \sin^2\theta}$$

Several other identities involving the del operator will come in handy in our study of optics:

$$\nabla \times \nabla V = 0, \tag{2A.10}$$

$$\nabla \cdot \nabla \times \mathbf{E} = 0, \tag{2A.11}$$

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}, \tag{2A.12}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{A} \cdot (\nabla \times \mathbf{B}) + \mathbf{B} \cdot (\nabla \times \mathbf{A}). \tag{2A.13}$$

We have displayed all of the above relations in a Cartesian coordinate system; similar expressions can be derived using spherical coordinates. We list them in Table 2A.1 for the reader's convenience.

Appendix 2B Electromagnetic Units

Of all topics in physics, perhaps the one that introduces the most confusion is the subject of electromagnetic units. In this book, we use MKS units, but the reader will find most older optics books will use c.g.s. units.

Table 2B.1 *Electromagnetic units*

c.g.s.	Units	MKS	Units
c	cm/s	$\frac{1}{\sqrt{\mu_0 \varepsilon_0}}$	m/s
$\mathbf{D}$	$12\pi \times 10^{-5}$ statcoulomb/cm^2	$\sqrt{4\pi\varepsilon_0}\ \mathbf{D}$	coulomb/m
$\mathbf{B}$	10^4 gauss	$\sqrt{\frac{4\pi}{\mu_0}}\ \mathbf{B}$	weber/m^2
$\mathbf{H}$	$4\pi \times 10^{-3}$ oersted	$\sqrt{4\pi\mu_0}\ \mathbf{H}$	amp-turn/m
$\mathbf{E}$	$\frac{1}{3} \times 10^{-4}$ statcoulomb/cm	$\sqrt{4\pi\varepsilon_0}\ \mathbf{E}$	volt/m
$\mathbf{J}$	3×10^5 statamp/cm^2	$\frac{\mathbf{J}}{\sqrt{4\pi\varepsilon_0}}$	amp/m^2
σ	9×10^9 s^{-1}	$\frac{\sigma}{4\pi\varepsilon_0}$	mho/m
ρ	3×10^3 statcoulomb/cm^3	$\frac{\rho}{\sqrt{4\pi\varepsilon_0}}$	coulomb/m^3
ε		$\frac{\varepsilon}{\varepsilon_0}$	
μ		$\frac{\mu}{\mu_0}$	

In rationalized MKS units, ε and μ in a vacuum have the following values

$$\begin{aligned}
\varepsilon_0 &= 8.8542 \times 10^{-12}\ \mathrm{C/(N \cdot m^2)} \quad (= \mathrm{F/m}) \\
&\approx \left(\frac{1}{36\pi}\right) \times 10^{-9}\ \mathrm{F/m}, \\
\mu_0 &= 4\pi \times 10^{-7}\mathrm{N \cdot s^2/C^2} \quad (= \mathrm{H/m}).
\end{aligned}$$

In c.g.s. units, $\varepsilon_0 = \mu_0 = 1$.

Maxwell's equations in c.g.s. units are written as

$$\begin{aligned}
\nabla \cdot \mathbf{D} &= 4\pi\rho, \\
\nabla \cdot \mathbf{B} &= 0, \\
\nabla \times \mathbf{H} &= \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}, \\
\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} &= 0.
\end{aligned}$$

A good discussion of the subject of units can be found in Jackson [1]. Here, we provide a brief table to aid the reader in converting from one system to the other.

REFERENCE

1. Jackson, J.D., *Clasical Electrodynamics*. 3rd ed. 1998: Wiley.

3

Reflection and Refraction

3.1 Introduction 52
3.2 Reflection and Transmission at a Discontinuity 53
3.3 Laws of Reflection and Refraction 56
3.4 Fresnel's Formula 58
3.5 Reflected and Transmitted Energy 63
3.6 Normal Incidence 65
3.7 Polarization by Reflection 68
3.8 Total Reflection 69
3.9 Reflection from a Conductor 73
3.10 Summary 75
3.11 Problems 77
References 78

3.1 Introduction

In Chapter 2, we treated the propagation of light in a uniform medium using Maxwell's equations. In this chapter, we wish to explore what happens to the propagation of a light wave when the electrical properties of the medium change in a discontinuous way. We will find that the wave will experience reflection at the boundary between two media with different electromagnetic properties. The light transmitted across the boundary will undergo a change in propagation direction. This direction change is called *refraction*. We will make the first of several derivations of the laws of refraction and reflection, here relying only on the wave properties of light to obtain these laws. To obtain the amplitude of the reflected and refracted light waves, we will use boundary conditions developed in classical electromagnetic theory for Maxwell's equations.

Once equations for the reflected and transmitted amplitudes have been obtained, we will consider light incident normal to the boundary to simplify the equations for the amplitudes of reflected and transmitted waves. With this simplification, it will become obvious that the fractions of the wave reflected and transmitted at the boundary between two media depend on the relative propagation velocities of the wave in the two media.

We will find that there is an angle, called Brewster's angle, for which light reflected from a boundary will be linearly polarized. There is also a set of conditions for which all light incident on a boundary will be reflected. A few of the properties of this reflected wave, called a totally reflected wave, will be discussed.

Use of reflection existed before written history, as is evidenced by the discovery of a mirror in Turkey from the period around 8000 years ago [1]. Some of the earliest written comments about reflection can be found in *Exodus* 38:8 and *Job* 37:18. **Euclid** in about 300 BC discussed the focus of a spherical mirror in his book *Catoptrics*. **Cleomedes** (AD 50) discussed refraction of light at an air–water interface. He described an experiment whereby a coin at the bottom of a bowl, and hidden by the bowl's sides, could be made visible by pouring water in the bowl.

Claudius Ptolemy of Alexandria (AD 139) made tables of the angles of incidence and refraction. His work is one of the few examples of experiment during that time. The concept of the sine of an angle was not yet developed, so his tables were only approximately correct. One of the most interesting individuals active in optics during the Middle Ages was **Alhazen** (Abû 'Alî al Hasan ibn al Hasan ibn Al Haitam, **965–1038**), who developed optics during the golden age of the Arabic empire. He added, to the law of reflection developed by Ptolemy, the fact that the incident angle and the reflected angle lie in the same plane, called the plane of incidence. He also corrected Ptolemy's tables of incident and refracted angles. Alhazen failed, however, to discover the law of refraction. (Alhazen was a successful optical scientist, but some of his civil engineering projects got him into trouble with his caliph, and Alhazen had to feign insanity and hide to escape the caliph's wrath.)

Vitello (in 1270) repeated Ptolemy's experiments, but also failed to discover the law of refraction. **Johannes Kepler (1571–1630)** gave a broad outline of the correct theory of the telescope and discussed total internal reflection without knowledge of the law of refraction. He used an empirical expression, $\theta_i = n\theta_r$, where $n = 3/2$.

Modern Optics. Second Edition. B. D. Guenther.

The law of refraction was discovered, evidently through experimentation, by **Willebrord Snellius (Willebrord Snel van Royen: 1580–1626)**, a professor of mathematics at Leiden. He never published, but Huygens and Isaak Voss claimed to have examined Snell's manuscript. Ibn Sahl is credited with the first discovery in 984, but he failed to get credit. **René Descartes (1596–1650)**, in 1637, deduced the law of refraction theoretically and expressed it in its present form. It is interesting that one of his assumptions was wrong. He often confidently deduced theory without allowing himself to be disturbed by any possible discrepancy between his final conclusions and the actual facts. Later, **Pierre de Fermat (1601–1665)** deduced the law of refraction from the assumption that light travels from a point in one medium to a point in another medium in the least time.

3.2 Reflection and Transmission at a Discontinuity

To develop the laws of reflection and refraction, we will first consider the one-dimensional problem of a vibrating string where the string has a discontinuous change in its mass/unit length at the origin of an (x, y) coordinate system (Figure 3.1). The string is stretched along the x-direction, and a wave propagating along the string in the positive x-direction creates a displacement in the y-direction. An example of such a string might be an "A" string of a guitar with part of its winding removed. In a string under tension T, the phase velocity of a wave is

$$\sqrt{\frac{T}{\mu}}. \tag{3.1}$$

[A dimensional analysis of this ratio,

$$T/\mu \Rightarrow (\text{kg} \cdot \text{m/s}^2)/(\text{kg/m}) = \text{m}^2/\text{s}^2 = (\text{velocity})^2,$$

supports the association of the ratio with velocity.] Therefore, a nonuniform wave propagation velocity will result from a nonuniform mass/unit length. What happens to the wave motion on a string with a nonuniform propagation velocity? Assume that the tension is the same throughout the string and that the mass/unit length is μ_1 for $x < 0$ and μ_2 for $x > 0$, as shown in Figure 3.1. We have two boundary conditions that must be satisfied at $x = 0$.

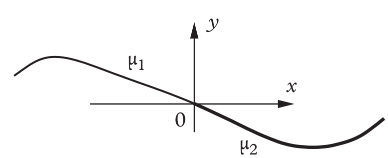

Figure 3.1 *String with a nonuniform mass/unit length. For $x < 0$, the mass/unit length is μ_1, while for $x > 0$, the mass/unit length is μ_2. The tension is the same throughout the string.*

1. The displacement at $x = 0$ must be the same for the two strings. If this were not the case, the string would be broken.
2. The slopes of the two strings must be the same at $x = 0$. We assume that the string undergoes small displacements so $\sin\theta \approx \tan\theta = dy/dx$. The vertical force at any point is therefore $T_y = T dy/dx$. If the slopes were not equal, there would be a finite vertical force acting on an infinitesimal mass, and the acceleration at $x = 0$ would be infinite.

Initially, we will assume that there are two waves present. One wave is incident from the left and propagating toward $x = 0$, traveling from left to right in Figure 3.1. This incident wave is given by

$$y_i = Y_i \cos(\omega_i t - k_i x), \qquad x < 0.$$

The speed of the wave is

$$c_1 = \sqrt{\frac{T}{\mu_1}}.$$

At the origin, $x = 0$, the wave and its first derivative are

$$y_i\Big|_{x=0} = Y_i \cos\omega_i t, \qquad \frac{dy_i}{dx}\Big|_{x=0} = k_i Y_i \sin\omega_i t.$$

The second wave, called the transmitted wave, is one that has propagated past $x = 0$, also traveling from left to right in Figure 3.1. The transmitted wave is represented by

$$y_t = Y_t \cos(\omega_t t - k_t x), \qquad x > 0,$$

with a propagation velocity given by $\{1\}$. At $x = 0$, the wave and its first derivative are

$$y_t\Big|_{x=0} = Y_t \cos\omega_t t, \qquad \frac{dy_t}{dx}\Big|_{x=0} = k_t Y_t \sin\omega_t t.$$

We will attempt to satisfy the boundary conditions with these two waves, but will discover that a third wave will be needed if the boundary conditions are to be satisfied.

The first boundary condition requires that the waves be equal at $x = 0$ and is met if $Y_i = Y_t$. Since the boundary condition must hold for all time, the frequencies of the waves propagating at the two different velocities must be the same, $\omega_i = \omega_t$.

The second boundary condition requires that the first derivatives of the two waves must be equal at $x = 0$, and we find again that $\omega_i = \omega_t$, since the equality must hold for all time. The second boundary condition also requires that

$$k_i Y_i = k_t Y_t,$$

but, because $Y_i = Y_t$ from the first boundary condition, we must have $k_i = k_t$. The definition (1.2) allows the replacement of the propagation constants by $\omega_i/c_1 = \omega_t/c_2$. Because the two frequencies are required to be equal by the boundary conditions, we are led to a contradiction of the initial assumption that the propagation velocities in the two dissimilar string segments are unequal, $c_1 \neq c_2$.

To overcome this contradiction and satisfy the boundary conditions, a reflected wave must be introduced, traveling from right to left and originating at $x = 0$. The reflected wave is defined as

$$y_r = Y_r \cos(\omega_r + k_r x), \qquad x < 0.$$

Its propagation velocity is c_1. The evaluation of the wave and its first derivative at the origin yields

$$y_r\Big|_{x=0} = Y_r \cos\omega_r t, \qquad \frac{dy_r}{dx}\Big|_{x=0} = -k_r Y_r \sin\omega_r t.$$

With the new wave, the boundary conditions are written

$$y_i + y_r - y_t\Big|_{x=0} = 0, \qquad \frac{dy_i}{dt} + \frac{dy_r}{dt} + \frac{dy_t}{dt}\Big|_{x=0} = 0.$$

These two equations are satisfied if

$$\omega_i = \omega_r = \omega_t, \tag{3.2}$$

$$Y_i + Y_r = Y_t, \tag{3.3}$$

$$k_i Y_i - k_r Y_r = k_t Y_t. \tag{3.4}$$

Solving (3.3) and (3.4) simultaneously yields

$$Y_t = \frac{k_i + k_r}{k_t + k_r} Y_i = \frac{\omega\left(\dfrac{1}{c_1} + \dfrac{1}{c_1}\right)}{\omega\left(\dfrac{1}{c_1} + \dfrac{1}{c_2}\right)} Y_i = \frac{2\dfrac{\omega}{c_1}}{\left(\dfrac{\omega}{c_1 c_2}\right)(c_1 + c_2)} Y_i.$$

The transmission coefficient at the junction of the two dissimilar strings is defined as $\tau = Y_t/Y_i$ and is given by

$$\tau = \frac{2c_2}{c_1 + c_2}. \tag{3.5}$$

In the same way, a reflection coefficient can also be defined at the junction where the mass/unit length of the string changes:

$$\rho = \frac{Y_r}{Y_i} = \frac{c_2 - c_1}{c_1 + c_2}. \tag{3.6}$$

If $c_1 = c_2$ then $Y_r = 0$, there is no reflected wave, and the incident and transmitted waves have equal amplitudes, $Y_i = Y_t$. The wave propagates pass the point $x = 0$ without change. Note that the transmission and reflection coefficients are functions only of the wave velocities. This same dependence upon propagation velocity will be found for light waves.

The frequency across the discontinuity is required, by the boundary conditions, to be a constant, but the velocity changes. This means that the wavelength must also change, as is shown in Figure 3.2. There are two waves to the left of the origin. From the principle of superposition, the displacement of the string at any time is a sum of these two waves: the incident wave and the reflected wave. The addition of the incident and reflected waves leads to a standing wave if the wave is totally reflected at the discontinuity.

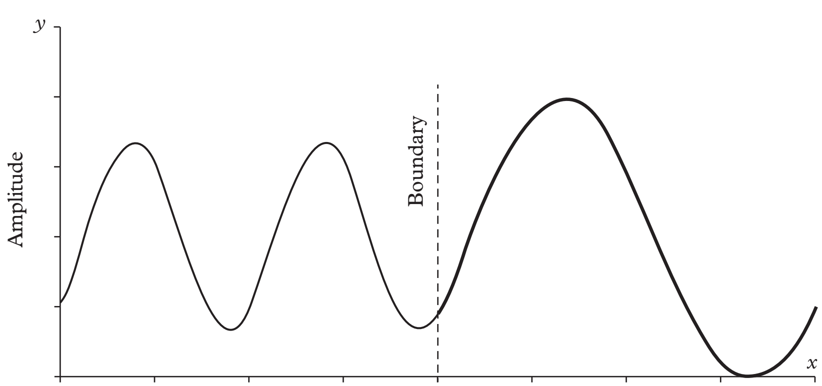

Figure 3.2 *Two strings of unequal mass/unit length joined at x = 0. The wave velocity of the string to the left is 20 m/s and that to the right is 10 m/s. The wave is incident onto the junction from the left and has amplitude of 3 cm and a wavelength of 1 m. The plot shows the resultant waves: on the right of x = 0 the transmitted wave and on the left of x = 0 the sum of the incident and reflected waves.*